

RESULT 4
US-09-509-814A-1
Sequence 1, Application US/09509814A
Patent No. 6376227
GENERAL INFORMATION:
APPLICANT: TAKAIWA, MIKIO
APPLICANT: OKUDA, MITSUYOSHI
APPLICANT: SAEKI, KATSUHIRO
APPLICANT: KUBOTA, HIROMI
APPLICANT: HITOMI, JUN
APPLICANT: KAGEYAMA, YASUSHI
APPLICANT: SHIKATA, SHITSUM
APPLICANT: NOMURA, MASAFUMI
TITLE OF INVENTION: ALKALINE PROTEASE
FILE REFERENCE: 0327-0832-0PCT
CURRENT APPLICATION NUMBER: US/09/509,814A
CURRENT FILING DATE: 2000-04-06
PRIOR APPLICATION NUMBER: PCT/JP99/04528
PRIOR FILING DATE: 1998-10-07
PRIOR APPLICATION NUMBER: JP 9-274570
PRIOR FILING DATE: 1997-06-08
NUMBER OF SEQ ID NOS: 24
SOFTWARE: PatentIn version 3.0
SEQ ID NO 1
LENGTH: 639
TYPE: PRT
ORGANISM: Bacillus sp.
FEATURE:
NAME/KEY: misc feature
LOCATION: (23)..(23)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (29)..(29)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (32)..(32)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (46)..(46)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (47)..(47)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (53)..(53)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (70)..(70)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (74)..(74)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (89)..(89)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (102)..(102)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (105)..(105)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (128)..(128)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (130)..(130)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (131)..(131)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature

LOCATION: (132)..(132)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (133)..(133)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (146)..(146)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (148)..(148)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (160)..(160)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (165)..(165)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (172)..(172)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (183)..(183)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (187)..(187)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (188)..(188)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (189)..(189)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (194)..(194)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (286)..(286)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (306)..(306)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (324)..(324)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (369)..(369)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (431)..(431)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (501)..(501)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (531)..(531)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (541)..(541)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (584)..(584)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (591)..(591)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (592)..(592)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (594)..(594)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (595)..(595)

LENGTH: 640
TYPE: prt
ORGANISM: Bacillus sp.
FEATURE:
NAME/KEY: misc_feature
LOCATION: (3)..(3)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (24)..(24)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (30)..(30)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (33)..(33)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (47)..(47)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (48)..(48)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (54)..(54)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (71)..(71)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (75)..(75)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (90)..(90)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (103)..(103)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (106)..(106)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (129)..(129)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (131)..(131)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (132)..(132)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (133)..(133)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (134)..(134)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (147)..(147)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (149)..(149)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (161)..(161)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (166)..(166)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (173)..(173)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc_feature
LOCATION: (184)..(184)
OTHER INFORMATION: Xaa is any amino acid

NAME/KEY: misc feature
LOCATION: (188)..(188)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (189)..(189)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (190)..(190)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (195)..(195)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (287)..(287)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (307)..(307)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (325)..(325)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (370)..(370)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (432)..(432)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (502)..(502)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (532)..(532)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (542)..(542)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (585)..(585)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (592)..(592)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (593)..(593)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (595)..(595)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (596)..(596)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (597)..(597)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (612)..(612)
OTHER INFORMATION: Xaa is any amino acid
NAME/KEY: misc feature
LOCATION: (633)..(633)
OTHER INFORMATION: Xaa is any amino acid
US-09-814A-2

Query Match 95.7%; Score 2150; DB 4; Length 640;
Best Local Similarity 96.1%; Pred. No. 1.8e-164;
Matches 417; Conservative 1; Mismatches 16; Indels 0; Gaps 0;

Qy 1 NDVARGIVKADVAQSSYGLYGGQGVAVADTGLDGTGRNDSMHEAFRGKITALVALGRIN 60
Db 207 NDVARGIVKADVAQSSYGLYGGQGVAVADTGLDGTGRNDSMHEAFRGKITALVALGRIN 266

Qy 61 NANTNNGHGHVAGSVLGNSTNKGMAPOANLVFQSIIMDSGGGLGLPLSNLQTLFQAVS 120
Db 267 NANTNNGHGHVAGSVLGNSTNKGMAPOANLVFQSIIMDSGGGLGLPLSNLQTLFQAVS 326

RESULT 6
US-08-873-479-42
; Sequence 42, Application US/08873479
; Patent No. 5891701
; GENERAL INFORMATION:
; APPLICANT: Sloma, Alan
; TITLE OF INVENTION: Nucleic Acids Encoding A Polypeptide
; TITLE OF INVENTION: Having Protease Activity
; NUMBER OF SEQUENCES: 57
; CORRESPONDENCE ADDRESS:
; ADDRESS: No. 5891701 No. 5891701disk of No. 5891701th America
; STREET: 405 Lexington Avenue
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10174
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/873,479
; FILING DATE: 12-JUN-1997
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Agis, Cheryl H
; REGISTRATION NUMBER: 34,086
; REFERENCE/DOCKET NUMBER: 5251.000-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; TELEX:
; INFORMATION FOR SEQ ID NO: 42:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 641 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
US-08-873-479-42

Query Match 94.4%; Score 2120.5; DB 2; Length 641;
Best Local Similarity 93.3%; Pred. No. 4.2e-162;
Matches 405; Conservative 20; Mismatches 8; Indels 1; Gaps 1;

```

QY 1 NDVARGIVKADVAQSSVGLYGQGOIVAVADTGLDTRNDSSMHEAFRGKITALYALGRTN 60
DB 209 NDVARGIVKADVAQNFGLYGQGOIVAVADTGLDTRNDSSMHEAFRGKITALYALGRTN 268
QY 61 NANTNGHGHVAGSVLNGSTNKGMAPQANLVFQSIMDSGGGLGGLPSNLQTLFSQAYS 120
DB 269 NANTDPNGHGHVAGSVLGN-ATNKGMAPQANLVFQSIMDSGGGLGGLPANLQTLFSQAYS 327
QY 121 AGARIHNSNGAAVNGAYTTDSNRVDDYVRKNDMTILFAAGNEGPNNGGTISAPGTAKNAI 180
DB 328 AGARIHNSNGAPVNGAYTTDSNRVDDYVRKNDMTILFAAGNEGPNNGGTISAPGTAKNAI 387
QY 181 TVGATENLRPSFGSYADNINHVAQFSSRGPTKDGRIKPDVMAPTFTILSARSLAPDSF 240
DB 388 TVGATENLRPSFGSYADNINHVAQFSSRGPTKDGRIKPDVMAPTFTILSARSLAPDSF 447
QY 241 WANHDSKYAVMGGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAAIAGAADIIGLY 300
DB 448 WANHDSKYAVMGGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAAIAGAADVGLGF 507
QY 301 PNGNQGWGRVTLDKSLNVAYVNESSLSSTSQKATYSFTATAGKPLKISLVMSDAPASTTA 360
DB 508 PNGNQGWGRVTLDKSLNVAYVNESSLSSTSQKATYSFTATAGKPLKISLVMSDAPASTTA 567
QY 361 SVTLVNDLNLVITAPNGTQVVGNDFTSPYNDNDWGRNNVNFINAPQSGTITIEVOAYN 420
DB 568 SVTLVNDLNLVITAPNGTQVVGNDFTSPYNDNDWGRNNVNFINAPQSGTITIEVOAYN 627
QY 421 VPVGPQTFFSLAIVN 434
DB 628 VPVGPQTFFSLAIVH 641

```

RESULT 7

```

US-09-104-623A-4
; Sequence 4, Application US/09104623A
; Patent No. 6303752
; GENERAL INFORMATION:
; APPLICANT: Olsen, Arne Agerlin
; APPLICANT: Fatum, Tine Muxoll
; APPLICANT: Deussen, Heinz-Josef
; APPLICANT: Roggen, Erwin Ludo
; TITLE OF INVENTION: A Modified Polypeptide
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6303752o No. 6303752disk of No. 6303752th America, Inc.
; STREET: 405 Lexington Avenue
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10174
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/104,623A
; FILING DATE: 25-JUN-1998
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Rozek, Carol
; REGISTRATION NUMBER: 36,993
; REFERENCE/DOCKET NUMBER: 5256.200-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 433 amino acids
; TYPE: amino acid
; STRANDEDNESS:

```

```

; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; STRAIN: Bacillus sp. Y
US-09-104-623A-4

```

```

Query Match 88.2%; Score 1981.5; DB 4; Length 433;
Best Local Similarity 87.1%; Pred. No. 3.6e-151;
Matches 378; Conservative 30; Mismatches 25; Indels 1; Gaps 1;

```

```

QY 1 NDVARGIVKADVAQSSVGLYGQGOIVAVADTGLDTRNDSSMHEAFRGKITALYALGRTN 60
DB 1 NDVARGIVKADVAQNFGLYGQGOIVAVADTGLDTRNDSSMHEAFRGKITALYALGRTN 60
QY 61 NANTNGHGHVAGSVLNGSTNKGMAPQANLVFQSIMDSGGGLGGLPSNLQTLFSQAYS 120
DB 61 NASDPNGHGHVAGSVLGN-ALNKGMAPQANLVFQSIMDSGGGLGGLPSNLQTLFSQAWN 119
QY 121 AGARIHNSNGAAVNGAYTTDSNRVDDYVRKNDMTILFAAGNEGPNNGGTISAPGTAKNAI 180
DB 120 AGARIHNSNGAPVNGAYTTANSQVDEYVRNNDMTVLFAAGNEGPNNGGTISAPGTAKNAI 179
QY 181 TVGATENLRPSFGSYADNINHVAQFSSRGPTKDGRIKPDVMAPTFTILSARSLAPDSF 240
DB 180 TVGATENLRPSFGSIADNPNHIAQFSSRGATRDGRIKPDVTAPGTFTILSARSLAPDSF 239
QY 241 WANHDSKYAVMGGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAAIAGAADIIGLY 300
DB 240 WANYSKYAVMGGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAAIAGATDVGLGY 299
QY 301 PNGNQGWGRVTLDKSLNVAYVNESSLSSTSQKATYSFTATAGKPLKISLVMSDAPASTTA 360
DB 300 PSGDQGWGRVTLDKSLNVAYVNEATATATQKATYSFQAQAGKPLKISLVMTDAPGSTTA 359
QY 361 SVTLVNDLNLVITAPNGTQVVGNDFTSPYNDNDWGRNNVNFINAPQSGTITIEVOAYN 420
DB 360 SYTLVNDLNLVITAPNGQKYVGNDFSYPYDNNWGRNNVNFINAPQSGTITIEVOAYN 419
QY 421 VPVGPQTFFSLAIVN 434
DB 420 VPVGPQTFFSLAIVH 433

```

RESULT 8

```

US-09-019-532-4
; Sequence 4, Application US/09019532B
; Patent No. 6416756
; GENERAL INFORMATION:
; APPLICANT: Olsen, Arne Agerlin
; APPLICANT: Prent, Annette
; TITLE OF INVENTION: A Modified Enzyme for Skin Care
; FILE REFERENCE: 4922.204-US
; CURRENT APPLICATION NUMBER: US/09/019,532B
; CURRENT FILING DATE: 1998-02-05
; EARLIER APPLICATION NUMBER: 0038/97
; EARLIER FILING DATE: 1997-01-10
; EARLIER APPLICATION NUMBER: 0754/97
; EARLIER FILING DATE: 1997-06-25
; EARLIER APPLICATION NUMBER: 60/051,381
; EARLIER FILING DATE: 1997-07-07
; EARLIER APPLICATION NUMBER: PCT/DK98/00015
; EARLIER FILING DATE: 1998-01-12
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 433
; TYPE: PRT
; ORGANISM: Bacillus sp.
US-09-019-532-4

```

```

Query Match 88.2%; Score 1981.5; DB 4; Length 433;
Best Local Similarity 87.1%; Pred. No. 3.6e-151;
Matches 378; Conservative 30; Mismatches 25; Indels 1; Gaps 1;

```


QY 1 NDVARGIVKADVAOSSYGLYGQGOI VAVADTGLDGTGRNDSMHEAFRGKITALYALGRTN 60
Db 1 NDVARGIVKADVAQNNGYGLYGQGOI VAVADTGLDGTGRNDSMHEAFRGKITALYALGRTN 60
QY 61 NANTNGHGHVAGSVLNGSTNKGMAPOANLVFQSIMDSGGGLGGLPSNLQTLFQAYS 120
Db 61 NASDPNGHGHVAGSVLNG -ALNKGMAPOANLVFQSIMDSGGGLGGLPSNLQTLFQAYS 119
QY 121 AGARIHTNSWGAAYNGAYTTDSRNVDYVKNMTILFAAGNEGPNNGTTISAPGTAKNAI 180
Db 120 AGARIHTNSWGAAYNGAYTANSROVDYVRNNDMTVLFAAGNEGPNNGTTISAPGTAKNAI 179
QY 181 TVGATENLRPFSGYADNINHVAQSSRGPTKDGRIKPDVMAPOGTFILSARSSSLAPDSSF 240
Db 180 TVGATENLRPFSGYADNINHVAQSSRGPTKDGRIKPDVMAPOGTFILSARSSSLAPDSSF 239
QY 241 WANHDSKYAYMGTSMAPIVAGNVAQLRHFVKRGITPKPSLLKAALIAAGADIGLY 300
Db 240 WANYNSKYAYMGTSMAPIVAGNVAQLRHFVKRGITPKPSLLKAALIAAGADIGLY 299
QY 301 PNGNGWGRVTLDSLVNVAYNNESSLSTOKATYSFTATAGKPLKISLVNSDAPASTTA 360
Db 300 PSQGQWGRVTLDSLVNVAYNNEATATGKATYSFOAQAGKPLKISLVNTDAPGSTTA 359
QY 361 SVTLVNDLNLVITAPNGTQYVGNDFTSYNDNDGRNNDVNFNAPQSGTYTIEVQAYN 420
Db 360 SYTLVNDLNLVITAPNGKQYVGNDFSYNDNDGRNNDVNFNAPQSGTYTIEVQAYN 419
QY 421 VPVGPQTFSLAIVN 434
Db 420 VPSGPQRFSLAIVH 433

RESULT 9

US-09-338-746-4

; Sequence 4, Application US/09338746

; Patent No. 6638526

; GENERAL INFORMATION:

; APPLICANT: Deussen, Heinz-Josef

; APPLICANT: Olsen, Arne A.

; APPLICANT: Fatum, Tine M.

; APPLICANT: Roggen, Erwin L.

; TITLE OF INVENTION: A Polypeptide-Polymer Conjugate

; FILE REFERENCE: 5619-200-US

; CURRENT APPLICATION NUMBER: US/09/338,746

; CURRENT FILING DATE: 1998-06-23

; EARLIER APPLICATION NUMBER: PA 1998 00809

; EARLIER FILING DATE: 1998-06-23

; EARLIER APPLICATION NUMBER: 60/091,461

; EARLIER FILING DATE: 1998-07-01

; NUMBER OF SEQ ID NOS: 6

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 4

; LENGTH: 433

; TYPE: PRT

; ORGANISM: Bacillus sp.

US-09-338-746-4

Query Match 88.2%; Score 1981.5; DB 4; Length 433;

Best Local Similarity 87.1%; Pred. No. 3.6e-151;

Matches 378; Conservative 30; Mismatches 25; Indels 1; Gaps 1;

QY 1 NDVARGIVKADVAOSSYGLYGQGOI VAVADTGLDGTGRNDSMHEAFRGKITALYALGRTN 60
Db 1 NDVARGIVKADVAQNNGYGLYGQGOI VAVADTGLDGTGRNDSMHEAFRGKITALYALGRTN 60
QY 61 NANTNGHGHVAGSVLNGSTNKGMAPOANLVFQSIMDSGGGLGGLPSNLQTLFQAYS 120
Db 61 NASDPNGHGHVAGSVLNG -ALNKGMAPOANLVFQSIMDSGGGLGGLPSNLQTLFQAYS 119
QY 121 AGARIHTNSWGAAYNGAYTTDSRNVDYVKNMTILFAAGNEGPNNGTTISAPGTAKNAI 180

Db 120 AGARIHTNSWGAAYNGAYTANSROVDYVRNNDMTVLFAAGNEGPNNGTTISAPGTAKNAI 179
QY 181 TVGATENLRPFSGYADNINHVAQSSRGPTKDGRIKPDVMAPOGTFILSARSSSLAPDSSF 240
Db 180 TVGATENLRPFSGYADNINHVAQSSRGPTKDGRIKPDVMAPOGTFILSARSSSLAPDSSF 239
QY 241 WANHDSKYAYMGTSMAPIVAGNVAQLRHFVKRGITPKPSLLKAALIAAGADIGLY 300
Db 240 WANYNSKYAYMGTSMAPIVAGNVAQLRHFVKRGITPKPSLLKAALIAAGADIGLY 299
QY 301 PNGNGWGRVTLDSLVNVAYNNESSLSTOKATYSFTATAGKPLKISLVNSDAPASTTA 360
Db 300 PSQGQWGRVTLDSLVNVAYNNEATATGKATYSFOAQAGKPLKISLVNTDAPGSTTA 359
QY 361 SVTLVNDLNLVITAPNGTQYVGNDFTSYNDNDGRNNDVNFNAPQSGTYTIEVQAYN 420
Db 360 SYTLVNDLNLVITAPNGKQYVGNDFSYNDNDGRNNDVNFNAPQSGTYTIEVQAYN 419
QY 421 VPVGPQTFSLAIVN 434
Db 420 VPSGPQRFSLAIVH 433

RESULT 10

US-08-873-479-43

; Sequence 43, Application US/08873479

; Patent No. 5891701

; GENERAL INFORMATION:

; APPLICANT: Sloma, Alan

; APPLICANT: Lynne, Christianson

; TITLE OF INVENTION: Nucleic Acids Encoding A Polypeptide

; TITLE OF INVENTION: Having Protease Activity

; NUMBER OF SEQUENCES: 57

; CORRESPONDENCE ADDRESS:

; ADDRESS: No. 5891701 No. 5891701disk of No. 5891701th America

; CITY: New York

; STATE: NY

; COUNTRY: USA

; ZIP: 10174

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; SOFTWARE: FastSeq for Windows Version 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/873,479

; FILING DATE: 12-JUN-1997

; CLASSIFICATION: 530

; ATTORNEY/AGENT INFORMATION:

; NAME: Agis, Cheryl H

; REGISTRATION NUMBER: 34,086

; REFERENCE/DOCKET NUMBER: 5251.000-US

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 212-867-0123

; TELEFAX: 212-878-9655

; TELEX:

; INFORMATION FOR SEQ ID NO: 43:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 635 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

US-08-873-479-43

Query Match 88.2%; Score 1981.5; DB 2; Length 635;

Best Local Similarity 87.1%; Pred. No. 6.1e-151;

Matches 378; Conservative 30; Mismatches 25; Indels 1; Gaps 1;

QY 1 NDVARGIVKADVAOSSYGLYGQGOI VAVADTGLDGTGRNDSMHEAFRGKITALYALGRTN 60
Db 203 NDVARGIVKADVAQNNGYGLYGQGOI VAVADTGLDGTGRNDSMHEAFRGKITALYALGRTN 262

QY 61 NNDTNGHGHVAGSVLNGSTNKGMAPOANLVFOSIMDSGGGLGGLPSNLQTLFSQAYS 120
Db 263 NASDPNGHGHVAGSVLNG-ALNKGMAPOANLVFOSIMDSGGGLGGLPSNLQTLFSQAWN 321
QY 121 AGARIHTNSGAAVNGAYTTDSRNVDDYVRKNDMTILFAAGNEGPNNGTISAPGTAKNAI 180
Db 322 AGARIHTNSGAAVNGAYTANSRQVDEYVRNNDMTILFAAGNEGPNNGTISAPGTAKNAI 381
QY 181 TVGATENLRPSFGSVADNINHVAFSSRGPTKDGRIKPDVMAFGTILSARSSSLAPDSSF 240
Db 382 TVGATENLRPSFGSVADNINHVAFSSRGPTKDGRIKPDVMAFGTILSARSSSLAPDSSF 441
QY 241 WANHDSKYAYMGGTSMATPIVAGNVAQLREHFVYKRGITPKPSLLKAALIAGAADIGLY 300
Db 442 WANHDSKYAYMGGTSMATPIVAGNVAQLREHFVYKRGITPKPSLLKAALIAGAADIGLY 501
QY 301 PNGNOGWGRVTLDKSLNV 318
Db 502 PSQDQGWGRVTLDKSLNV 345
QY 361 SVTLVNDLNLVITAPNGTQVGNDFTPYNDNWDGRNNVNFVINAPOSQGTIIEVQAYN 420
Db 562 SVTLVNDLNLVITAPNGQYVGNDFTPYNDNWDGRNNVNFVINAPOSQGTIIEVQAYN 621
QY 421 VPVGPTQPSLAVN 434
Db 622 VPSPGQRFSLAVH 635

RESULT 11
US-09-512-251A-10
; Sequence 10, Application US/09512251A
; Patent No. 6555355
; GENERAL INFORMATION:
; APPLICANT: Hansen, Peter
; APPLICANT: Bauditz, Peter
; APPLICANT: Mikkelsen, Frank
; APPLICANT: Andersen, Kim
; TITLE OF INVENTION: Protease Variants and Compositions
; FILE REFERENCE: 5349.204-US
; CURRENT APPLICATION NUMBER: US/09/512,251A
; CURRENT FILING DATE: 2000-02-24
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 10
; LENGTH: 345
; TYPE: PRT
; ORGANISM: Bacillus
US-09-512-251A-10

Query Match 70.4%; Score 1581.5; DB 4; Length 345;
Best Local Similarity 95.3%; Pred. No. 3.4e-119;
Matches 303; Conservative 10; Mismatches 4; Indels 1; Gaps 1;
QY 1 NDVARGIVKADVAGSVLNGSTNKGMAPOANLVFOSIMDSGGGLGGLPSNLQTLFSQAYS 60
Db 29 NDVARGIVKADVAGSVLNGSTNKGMAPOANLVFOSIMDSGGGLGGLPSNLQTLFSQAYS 88
QY 61 NNDTNGHGHVAGSVLNGSTNKGMAPOANLVFOSIMDSGGGLGGLPSNLQTLFSQAYS 120
Db 89 NNDTNGHGHVAGSVLNG-ATNKGMAPOANLVFOSIMDSGGGLGGLPSNLQTLFSQAYS 147
QY 121 AGARIHTNSGAAVNGAYTTDSRNVDDYVRKNDMTILFAAGNEGPNNGTISAPGTAKNAI 180
Db 148 AGARIHTNSGAAVNGAYTTDSRNVDDYVRKNDMTILFAAGNEGPNNGTISAPGTAKNAI 207
QY 181 TVGATENLRPSFGSVADNINHVAFSSRGPTKDGRIKPDVMAFGTILSARSSSLAPDSSF 240
Db 208 TVGATENLRPSFGSVADNINHVAFSSRGPTKDGRIKPDVMAFGTILSARSSSLAPDSSF 267
QY 241 WANHDSKYAYMGGTSMATPIVAGNVAQLREHFVYKRGITPKPSLLKAALIAGAADIGLY 300
Db 268 WANHDSKYAYMGGTSMATPIVAGNVAQLREHFVYKRGITPKPSLLKAALIAGAADIGLY 327

QY 301 PNGNOGWGRVTLDKSLNV 318
Db 328 PNGNOGWGRVTLDKSLNV 345

RESULT 12
US-09-515-150A-10
; Sequence 10, Application US/09515150A
; Patent No. 6558938
; GENERAL INFORMATION:
; APPLICANT: Hansen, Peter
; APPLICANT: Bauditz, Peter
; APPLICANT: Mikkelsen, Frank
; APPLICANT: Andersen, Kim
; TITLE OF INVENTION: Protease Variants and Compositions
; FILE REFERENCE: 5348.204-US
; CURRENT APPLICATION NUMBER: US/09/515,150A
; CURRENT FILING DATE: 2000-02-29
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 10
; LENGTH: 345
; TYPE: PRT
; ORGANISM: Bacillus
US-09-515-150A-10

Query Match 70.4%; Score 1581.5; DB 4; Length 345;
Best Local Similarity 95.3%; Pred. No. 3.4e-119;
Matches 303; Conservative 10; Mismatches 4; Indels 1; Gaps 1;
QY 1 NDVARGIVKADVAGSVLNGSTNKGMAPOANLVFOSIMDSGGGLGGLPSNLQTLFSQAYS 60
Db 29 NDVARGIVKADVAGSVLNGSTNKGMAPOANLVFOSIMDSGGGLGGLPSNLQTLFSQAYS 88
QY 61 NNDTNGHGHVAGSVLNGSTNKGMAPOANLVFOSIMDSGGGLGGLPSNLQTLFSQAYS 120
Db 89 NNDTNGHGHVAGSVLNG-ATNKGMAPOANLVFOSIMDSGGGLGGLPSNLQTLFSQAYS 147
QY 121 AGARIHTNSGAAVNGAYTTDSRNVDDYVRKNDMTILFAAGNEGPNNGTISAPGTAKNAI 180
Db 148 AGARIHTNSGAAVNGAYTTDSRNVDDYVRKNDMTILFAAGNEGPNNGTISAPGTAKNAI 207
QY 181 TVGATENLRPSFGSVADNINHVAFSSRGPTKDGRIKPDVMAFGTILSARSSSLAPDSSF 240
Db 208 TVGATENLRPSFGSVADNINHVAFSSRGPTKDGRIKPDVMAFGTILSARSSSLAPDSSF 267
QY 241 WANHDSKYAYMGGTSMATPIVAGNVAQLREHFVYKRGITPKPSLLKAALIAGAADIGLY 300
Db 268 WANHDSKYAYMGGTSMATPIVAGNVAQLREHFVYKRGITPKPSLLKAALIAGAADIGLY 327

RESULT 13
US-09-196-281-13
; Sequence 13, Application US/09196281A
; Patent No. 6605458
; GENERAL INFORMATION:
; APPLICANT: Hansen, Peter K.
; APPLICANT: Bauditz, Peter
; APPLICANT: Mikkelsen, Frank
; TITLE OF INVENTION: Protease Variants And Compositions
; FILE REFERENCE: 5435.200-US
; CURRENT APPLICATION NUMBER: US/09/196,281A
; CURRENT FILING DATE: 1998-11-19
; EARLIER APPLICATION NUMBER: 1332/97
; EARLIER FILING DATE: 1997-11-21
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 13

```

; LENGTH: 345
; TYPE: PRT
; ORGANISM: Bacillus
US-09-196-281-13

Query Match 70.4%; Score 1581.5; DB 4; Length 345;
Best Local Similarity 95.3%; Pred. No. 3.4e-119;
Matches 303; Conservative 10; Mismatches 4; Indels 1; Gaps 1;

QY 1 NDVARGIVKADVAQSSVGLYQGQGI VAVADTGLDTRNDSSMHAFRGKITALYALGRTN 60
Db 29 NDVARGIVKADVAQNGLYQGQGI VAVADTGLDTRNDSSMHAFRGKITALYALGRTN 88
QY 61 NANDTNGHGTAVAGSVLNGSTKNGMAPOANLVFQSTMDSGGLGGGLPSNLQTLFSAYS 120
Db 89 NANDPNHGHTAVAGSVLGN-ATNKGMAPOANLVFQSTMDSGGLGGGLPANLQTLFSAYS 147
QY 121 AGARIHTNSWGAAVNGAYTTTDSNRVDDYVRKNDMTILFAAGNEGPGGTTISAPGTAKNAI 180
Db 148 AGARIHTNSGAPVNGAYTTTDSNRVDDYVRKNDMTILFAAGNEGPGGTTISAPGTAKNAI 207
QY 181 TVGATENLRSFGSYADNINHVAFSSRGPTKGRIPKDVWAPQTFILSARSSLAPDSSF 240
Db 208 TVGATENLRSFGSYADNINHVAFSSRGPTKGRIPKDVWAPQTFILSARSSLAPDSSF 267
QY 241 WANHDSKYAYNGGTSMATPIVAGNVAQLREHFVNVRGTTPKPSLLKAALJAGAADIGLGY 300
Db 268 WANHDSKYAYNGGTSMATPIVAGNVAQLREHFVNVRGTTPKPSLLKAALJAGAADVGLGF 327
QY 301 PNGNQGWGRVTLDKSLNV 318
Db 328 PNGNQGWGRVTLDKSLNV 345

RESULT 14
US-08-894-818B-1
; Sequence 1, Application US/08894818B
; Patent No. 6261822
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: YAMAMOTO, Katsuhiko
; APPLICANT: MITTA, Masanori
; APPLICANT: ASADA, Kiyozo
; APPLICANT: TSUNASAWA, Susumu
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: HYPERTHERMOSTABLE PROTEASE GENES
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Browdy and Neimark
; STREET: 419 Seventh Street N.W., Ste. 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: United States of America
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/894,818B
; FILING DATE: 20-MAY-1998
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/JP96/03253
; FILING DATE: 07-NOV-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 323285/1995
; FILING DATE: 12-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Browdy, Roger L.
; REGISTRATION NUMBER: 25,618

```

Search completed: March 10, 2004, 14:55:45
Job time : 23 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2004, 14:54:44 ; Search time 40 Seconds
(without alignments)
2291.013 Million cell updates/sec

Title: US-09-985-689A-1-COPY

Perfect score: 2247

Sequence: 1 NDVARGIVKADVAQSSVGLY.....EVQAYNVPGPQFFSLAIVN 434

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 809742 seqs, 211153259 residues

Total number of hits satisfying chosen parameters: 809742

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|--------|-------------|--------|----|------------------|
| 1 | 2242 | 99.8 | 434 | 10 | US-09-985-689A-1 |
| 2 | 2242 | 99.8 | 434 | 15 | US-10-385-662-2 |
| 3 | 2186 | 97.3 | 434 | 10 | US-09-985-689A-2 |
| 4 | 2138 | 95.1 | 434 | 10 | US-09-985-689A-6 |
| 5 | 2120.5 | 94.4 | 433 | 10 | US-09-985-689A-7 |
| 6 | 1993.5 | 88.7 | 433 | 10 | US-09-985-689A-5 |
| 7 | 1989.5 | 88.5 | 433 | 10 | US-09-985-689A-3 |
| 8 | 1982.5 | 88.2 | 433 | 10 | US-09-985-689A-4 |
| 9 | 1581.5 | 70.4 | 345 | 14 | US-10-336-324-10 |
| 10 | 1581.5 | 70.4 | 345 | 14 | US-10-403-105-13 |
| 11 | 447.5 | 19.9 | 659 | 13 | US-10-090-624-12 |
| 12 | 409 | 18.2 | 412 | 13 | US-10-090-624-1 |
| 13 | 409 | 18.2 | 522 | 13 | US-10-090-624-4 |
| 14 | 409 | 18.2 | 654 | 13 | US-10-090-624-16 |
| 15 | 366 | 16.3 | 1079 | 14 | US-10-112-488-39 |

| | | | | | | |
|----|-------|------|-------|----|---------------------|--------------------|
| 16 | 363.5 | 16.2 | 1208 | 14 | US-10-156-761-13251 | Sequence 13251, A |
| 17 | 348 | 15.5 | 1139 | 14 | US-10-156-761-10856 | Sequence 10856, A |
| 18 | 336 | 15.0 | 1237 | 14 | US-10-314-657-4 | Sequence 4, Appli |
| 19 | 306.5 | 13.6 | 519 | 15 | US-10-084-846A-114 | Sequence 114, App |
| 20 | 306.5 | 13.6 | 19725 | 15 | US-10-084-846A-4 | Sequence 4, Appli |
| 21 | 305.5 | 13.6 | 1398 | 13 | US-10-090-624-6 | Sequence 6, Appli |
| 22 | 305.5 | 12.5 | 595 | 10 | US-09-927-827-59 | Sequence 59, Appli |
| 23 | 280.5 | 12.4 | 580 | 10 | US-09-927-827-59 | Sequence 55, Appli |
| 24 | 278 | 12.3 | 418 | 9 | US-09-966-921A-2 | Sequence 2, Appli |
| 25 | 275 | 12.3 | 1101 | 14 | US-10-156-761-12934 | Sequence 12934, A |
| 26 | 251 | 11.2 | 280 | 14 | US-10-209-812-2 | Sequence 2, Appli |
| 27 | 251 | 11.2 | 397 | 10 | US-09-779-334A-5 | Sequence 5, Appli |
| 28 | 247 | 11.0 | 271 | 10 | US-09-813-408-2 | Sequence 2, Appli |
| 29 | 246.5 | 11.0 | 379 | 10 | US-09-813-408-6 | Sequence 6, Appli |
| 30 | 246 | 10.9 | 271 | 14 | US-10-242-549-56 | Sequence 56, Appli |
| 31 | 244 | 10.9 | 271 | 14 | US-10-242-549-46 | Sequence 46, Appli |
| 32 | 244 | 10.9 | 271 | 14 | US-10-242-549-48 | Sequence 48, Appli |
| 33 | 244 | 10.9 | 271 | 14 | US-10-242-549-50 | Sequence 50, Appli |
| 34 | 244 | 10.9 | 271 | 14 | US-10-242-549-52 | Sequence 52, Appli |
| 35 | 244 | 10.9 | 271 | 14 | US-10-242-549-54 | Sequence 54, Appli |
| 36 | 244 | 10.9 | 271 | 14 | US-10-242-549-60 | Sequence 60, Appli |
| 37 | 242.5 | 10.8 | 627 | 10 | US-09-927-827-60 | Sequence 60, Appli |
| 38 | 242 | 10.8 | 271 | 14 | US-10-242-549-44 | Sequence 44, Appli |
| 39 | 242 | 10.8 | 271 | 14 | US-10-242-549-58 | Sequence 58, Appli |
| 40 | 240 | 10.7 | 269 | 8 | US-08-322-678-10 | Sequence 10, Appli |
| 41 | 240 | 10.7 | 269 | 9 | US-09-837-235-16 | Sequence 16, Appli |
| 42 | 240 | 10.7 | 269 | 9 | US-09-060-854B-6 | Sequence 6, Appli |
| 43 | 240 | 10.7 | 269 | 9 | US-09-975-139-1 | Sequence 1, Appli |
| 44 | 240 | 10.7 | 269 | 9 | US-09-976-414-8 | Sequence 8, Appli |
| 45 | 240 | 10.7 | 269 | 10 | US-09-736-116-49 | Sequence 49, Appli |

ALIGNMENTS

RESULT 1

US-09-985-689A-1

; Sequence 1, Application US/09985689A
; Publication No. US20030022351A1
; GENERAL INFORMATION:
; APPLICANT: HATADA, YUJI
; APPLICANT: OGAWA, AKINORI
; APPLICANT: KAGEYAMA, YASUSHI
; APPLICANT: SATO, TSUYOSHI
; APPLICANT: ARAKI, HIROYUKI
; APPLICANT: SUMITOMO, NOBUYUKI
; APPLICANT: OKUDA, MITSUYOSHI
; APPLICANT: SAEKI, KATSUHIKA
; TITLE OF INVENTION: Alkaline proteases
; FILE REFERENCE: 215483USO
; CURRENT APPLICATION NUMBER: US/09/985,689A
; PRIOR FILING DATE: 2002-07-01
; PRIOR APPLICATION NUMBER: JP P2000-355166
; PRIOR FILING DATE: 2000-11-22
; PRIOR APPLICATION NUMBER: JP P2001-114048
; PRIOR FILING DATE: 2001-04-12
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 434
; TYPE: PRT
; ORGANISM: Bacillus sp.
US-09-985-689A-1

Query Match 99.8%; Score 2242; DB 10; Length 434;

Best Local Similarity 99.8%; Pred. No. 3.9e-187; Indels 0; Gaps 0;

Matches 433; Conservative 1; Mismatches 0;

QY 1 NDVARGIVKADVAQSSVGLYGGQGIIVAVADTGLDTRNDSSMHEAFRCKITALYALGRTN 60

Db 1 NDVARGIVKADVAQSSVGLYGGQGIIVAVADTGLDTRNDSSMHEAFRCKITALYALGRTN 60

QY 61 NANDINGHGTAVAGSVLGNSTNKMAQANLVFQSIMDSGGGLGLSLNLTLPQAYS 120

Db 61 NANTNGHGHVAGSVLGNSTKGMAPQANLVFQSIMDSGGGLGGLPSNLQTLFSQAYS 120
 QY 121 AGARIHTNSGAAVNGAYTTDSRVDDYVRKNDMTILFAAGNEGPNNGGTISAPGTAKNAI 180
 Db 121 AGARIHTNSGAAVNGAYTTDSRVDDYVRKNDMTILFAAGNEGPNNGGTISAPGTAKNAI 180
 QY 181 TVGATENLRPSFGSYADNINHVAQFSSRGPTKDGRIKPDVMAFGTIFLSARSSLAPDSSF 240
 Db 181 TVGATENLRPSFGSYADNINHVAQFSSRGPTKDGRIKPDVMAFGTIFLSARSSLAPDSSF 240
 QY 241 WANEDSKYAYMGGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAGAADIIGLY 300
 Db 241 WANEDSKYAYMGGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAGAADIIGLY 300
 QY 301 PNGNQGWRTLDKSLNVAVYNESSLSSTSKATYSFTATAGKPLKISLWSDAPASTTA 360
 Db 301 PNGNQGWRTLDKSLNVAVYNESSLSSTSKATYSFTATAGKPLKISLWSDAPASTTA 360
 QY 361 SVTLVNDLNLVITAPNGTQYVGNDFTSYNDNDWGRNNVNFVINAPOSQGTYYIEVQAYN 420
 Db 361 SVTLVNDLNLVITAPNGTQYVGNDFTSYNDNDWGRNNVNFVINAPOSQGTYYIEVQAYN 420
 QY 421 VPVGPQTFFSLAIVN 434
 Db 421 VPVGPQTFFSLAIVN 434

RESULT 2

US-10-385-662-2
 ; Sequence 2, Application US/10385662
 ; Publication No. US2004002432A1
 ; GENERAL INFORMATION:
 ; APPLICANT: OKUDA, MITSUYOSHI
 ; APPLICANT: SATO, TSUYOSHI
 ; APPLICANT: SAITO, KAZUHIRO
 ; APPLICANT: SUMITOMO, NOBUYUKI
 ; APPLICANT: IZAWA, YOSHIFUMI
 ; APPLICANT: SAEKI, KATSUHIISA
 ; APPLICANT: KOBAYASHI, TOHRU
 ; APPLICANT: NOMURA, MASAFUMI
 ; TITLE OF INVENTION: Alkaline protease
 ; FILE REFERENCE: 234938USO
 ; CURRENT APPLICATION NUMBER: US/10385,662
 ; PRIOR FILING DATE: 2003-03-12
 ; PRIOR APPLICATION NUMBER: JP 2002-081428
 ; PRIOR FILING DATE: 2002-03-22
 ; PRIOR APPLICATION NUMBER: JP 2002-165987
 ; PRIOR FILING DATE: 2002-06-06
 ; PRIOR APPLICATION NUMBER: JP 2002-304230
 ; PRIOR FILING DATE: 2002-10-18
 ; PRIOR APPLICATION NUMBER: JP 2002-304231
 ; PRIOR FILING DATE: 2002-10-18
 ; NUMBER OF SEQ ID NOS: 20
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 2
 ; LENGTH: 434
 ; TYPE: PRT
 ; ORGANISM: Bacillus sp. KSM-KP43
 US-10-385-662-2

Query Match 99.8%; Score 2242; DB 15; Length 434;
 Best Local Similarity 99.8%; Pred. No. 3.9e-187;
 Matches 433; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 1 NDVARGIVKADVAQSSYGLYGQGIIVAVADTGLDTRNDSSMHEAFRGKITALYALGRTN 60
 Db 1 NDVARGIVKADVAQSSYGLYGQGIIVAVADTGLDTRNDSSMHEAFRGKITALYALGRTN 60
 QY 61 NANTNGHGHVAGSVLGNSTKGMAPQANLVFQSIMDSGGGLGGLPSNLQTLFSQAYS 120
 Db 61 NANTNGHGHVAGSVLGNSTKGMAPQANLVFQSIMDSGGGLGGLPSNLQTLFSQAYS 120

QY 121 AGARIHTNSGAAVNGAYTTDSRVDDYVRKNDMTILFAAGNEGPNNGGTISAPGTAKNAI 180
 Db 121 AGARIHTNSGAAVNGAYTTDSRVDDYVRKNDMTILFAAGNEGPNNGGTISAPGTAKNAI 180
 QY 181 TVGATENLRPSFGSYADNINHVAQFSSRGPTKDGRIKPDVMAFGTIFLSARSSLAPDSSF 240
 Db 181 TVGATENLRPSFGSYADNINHVAQFSSRGPTKDGRIKPDVMAFGTIFLSARSSLAPDSSF 240
 QY 241 WANEDSKYAYMGGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAGAADIIGLY 300
 Db 241 WANEDSKYAYMGGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAGAADIIGLY 300
 QY 301 PNGNQGWRTLDKSLNVAVYNESSLSSTSKATYSFTATAGKPLKISLWSDAPASTTA 360
 Db 301 PNGNQGWRTLDKSLNVAVYNESSLSSTSKATYSFTATAGKPLKISLWSDAPASTTA 360
 QY 361 SVTLVNDLNLVITAPNGTQYVGNDFTSYNDNDWGRNNVNFVINAPOSQGTYYIEVQAYN 420
 Db 361 SVTLVNDLNLVITAPNGTQYVGNDFTSYNDNDWGRNNVNFVINAPOSQGTYYIEVQAYN 420
 QY 421 VPVGPQTFFSLAIVN 434
 Db 421 VPVGPQTFFSLAIVN 434

RESULT 3

US-09-985-689A-2
 ; Sequence 2, Application US/09985689A
 ; Publication No. US20030022351A1
 ; GENERAL INFORMATION:
 ; APPLICANT: HATADA, YUJI
 ; APPLICANT: KAGEYAMA, YASUSHI
 ; APPLICANT: OGAWA, AKINORI
 ; APPLICANT: SATO, TSUYOSHI
 ; APPLICANT: ARAKI, HIROYUKI
 ; APPLICANT: SUMITOMO, NOBUYUKI
 ; APPLICANT: OKUDA, MITSUYOSHI
 ; APPLICANT: SAEKI, KATSUHIISA
 ; TITLE OF INVENTION: Alkaline proteases
 ; FILE REFERENCE: 215483USO
 ; CURRENT APPLICATION NUMBER: US/09/985,689A
 ; CURRENT FILING DATE: 2002-07-01
 ; PRIOR APPLICATION NUMBER: JP P2000-355166
 ; PRIOR FILING DATE: 2000-11-22
 ; PRIOR APPLICATION NUMBER: JP P2001-114048
 ; PRIOR FILING DATE: 2001-04-12
 ; NUMBER OF SEQ ID NOS: 7
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 2
 ; LENGTH: 434
 ; TYPE: PRT
 ; ORGANISM: Bacillus sp.
 US-09-985-689A-2

Query Match 97.3%; Score 2186; DB 10; Length 434;
 Best Local Similarity 96.3%; Pred. No. 3e-182;
 Matches 418; Conservative 14; Mismatches 2; Indels 0; Gaps 0;
 QY 1 NDVARGIVKADVAQSSYGLYGQGIIVAVADTGLDTRNDSSMHEAFRGKITALYALGRTN 60
 Db 1 NDVARGIVKADVAQSSYGLYGQGIIVAVADTGLDTRNDSSMHEAFRGKITALYALGRTN 60
 QY 61 NANTNGHGHVAGSVLGNSTKGMAPQANLVFQSIMDSGGGLGGLPSNLQTLFSQAYS 120
 Db 61 NANTNGHGHVAGSVLGNSTKGMAPQANLVFQSIMDSGGGLGGLPSNLQTLFSQAYS 120
 QY 121 AGARIHTNSGAAVNGAYTTDSRVDDYVRKNDMTILFAAGNEGPNNGGTISAPGTAKNAI 180
 Db 121 AGARIHTNSGAAVNGAYTTDSRVDDYVRKNDMTILFAAGNEGPNNGGTISAPGTAKNAI 180
 QY 181 TVGATENLRPSFGSYADNINHVAQFSSRGPTKDGRIKPDVMAFGTIFLSARSSLAPDSSF 240
 Db 181 TVGATENLRPSFGSYADNINHVAQFSSRGPTKDGRIKPDVMAFGTIFLSARSSLAPDSSF 240

Db 420 VPVSPQTFSLAIHV 433
|||||

RESULT 6

US-09-985-689A-5

; Sequence 5, Application US/09985689A

; Publication No. US20030022351A1

; GENERAL INFORMATION:

; APPLICANT: HATADA, YUJI

; APPLICANT: OGAWA, AKINORI

; APPLICANT: KAGEYAMA, YASUSHI

; APPLICANT: SATO, TSUYOSHI

; APPLICANT: ARAKI, HIROYUKI

; APPLICANT: SUMITOMO, NOBUYUKI

; APPLICANT: OKUDA, MITSUYOSHI

; APPLICANT: SAEKI, KATSUHISA

; TITLE OF INVENTION: Alkaline proteases

; FILE REFERENCE: 215483US0

; CURRENT APPLICATION NUMBER: US/09/985,689A

; CURRENT FILING DATE: 2002-07-01

; PRIOR APPLICATION NUMBER: JP P2000-355166

; PRIOR FILING DATE: 2000-11-22

; PRIOR APPLICATION NUMBER: JP P2001-114048

; PRIOR FILING DATE: 2001-04-12

; NUMBER OF SEQ ID NOS: 7

; SOFTWARE: Patentin version 3.1

; SEQ ID NO 5

; LENGTH: 433

; TYPE: PRT

; ORGANISM: Bacillus sp.

US-09-985-689A-5

Query Match 88.7%; Score 1993.5; DB 10; Length 433;

Best Local Similarity 87.6%; Pred. No. 1.9e-165; Indels 1; Gaps 1;

Matches 380; Conservative 29; Mismatches 24;

Qy 1 NDVARGIVKADVAQSSYGLYGQGVAVADTGLDGRNDSSMHEAFRGKITAYALGRTN 60
Db 1 NDVARGIVKADVAQNNYGLYGQGVAVADTGLDGRNDSSMHEAFRGKITAYALGRTN 60

Qy 61 NANDTNGHGHVAGSVLNGSTNKGMAPOANLVFQSIMDSGGGLGGLPSNLQTLFSQAYS 120
Db 61 NANDPNGHGHVAGSVLGN-ALNKGMAPOANLVFQSIMDSGGGLGGLPSNLQTLFSQAWN 119

Qy 121 AGARIHTNSWGAANGAYTTDSRNVDYVRKNDMTILFAAGNEGPNCGTISAPGTAKNAI 180
Db 120 AGARIHTNSWGAAPVNGAYTANSRQVDEYVRNNDMTVLFAAGNEGPNCGTISAPGTAKNAI 179

Qy 181 TVGATENLRPSFGSYADNINHVAQFSSRGTQDGRKPDVMAPTFILSARSSSLAPDSSF 240
Db 180 TVGATENYRPSFGSLADNPNHIAQFSSRGATRDGRIKPDVTAPGTFILSARSSSLAPDSSF 239

Qy 241 WANHDSKYAMGTSMTATPIVAGNVAQLREHFVKNRGITPKPSLLKAAALIAAGADIGLY 300
Db 240 WANYSKYAYMGTSMTATPIVAGNVAQLREHFVKNRGITPKPSLLKAAALIAAGADIGLY 299

Qy 301 PNGNQGWGRVTLDKSLNVAYNVESLSSTSQKATYSFTATAGKPLKISLWSDAPASTTA 360
Db 300 PSGDQGWGRVTLDKSLNVAYNVEATATGQKATYSFQAQKPLKISLWTDAPGSTTA 359

Qy 361 SVTLVNDLNLVITAPNGTQVGNDFSPYNDNWDGRNNVNFVINAPOSQGTITIEVQAYN 420
Db 360 SVTLVNDLNLVITAPNGQKIVGNDFSPYDNNWDGRNNVNFVINAPOSQGTITIEVQAYN 419

Qy 421 VPVGPQTFSLAIYN 434
Db 420 VPSPQRFSLAIHV 433

RESULT 7

US-09-985-689A-3

; Sequence 3, Application US/09985689A

; Publication No. US20030022351A1
; GENERAL INFORMATION:
; APPLICANT: HATADA, YUJI
; APPLICANT: OGAWA, AKINORI
; APPLICANT: KAGEYAMA, YASUSHI
; APPLICANT: SATO, TSUYOSHI
; APPLICANT: ARAKI, HIROYUKI
; APPLICANT: SUMITOMO, NOBUYUKI
; APPLICANT: OKUDA, MITSUYOSHI
; APPLICANT: SAEKI, KATSUHISA
; TITLE OF INVENTION: Alkaline proteases
; FILE REFERENCE: 215483US0
; CURRENT APPLICATION NUMBER: US/09/985,689A
; CURRENT FILING DATE: 2002-07-01
; PRIOR APPLICATION NUMBER: JP P2000-355166
; PRIOR FILING DATE: 2000-11-22
; PRIOR APPLICATION NUMBER: JP P2001-114048
; PRIOR FILING DATE: 2001-04-12
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 3
; LENGTH: 433
; TYPE: PRT
; ORGANISM: Bacillus sp.
US-09-985-689A-3

Query Match 88.5%; Score 1989.5; DB 10; Length 433;

Best Local Similarity 87.3%; Pred. No. 4.3e-165; Indels 1; Gaps 1;

Matches 379; Conservative 29; Mismatches 25;

Qy 1 NDVARGIVKADVAQSSYGLYGQGVAVADTGLDGRNDSSMHEAFRGKITAYALGRTN 60
Db 1 NDVARGIVKADVAQNNYGLYGQGVAVADTGLDGRNDSSMHEAFRGKITAYALGRTN 60

Qy 61 NANDTNGHGHVAGSVLNGSTNKGMAPOANLVFQSIMDSGGGLGGLPSNLQTLFSQAYS 120
Db 61 NANDPNGHGHVAGSVLGN-ALNKGMAPOANLVFQSIMDSGGGLGGLPSNLQTLFSQAWN 119

Qy 121 AGARIHTNSWGAANGAYTTDSRNVDYVRKNDMTILFAAGNEGPNCGTISAPGTAKNAI 180
Db 120 AGARIHTNSWGAAPVNGAYTANSRQVDEYVRNNDMTVLFAAGNEGPNCGTISAPGTAKNAI 179

Qy 181 TVGATENLRPSFGSYADNINHVAQFSSRGTQDGRKPDVMAPTFILSARSSSLAPDSSF 240
Db 180 TVGATENYRPSFGSLADNPNHIAQFSSRGATRDGRIKPDVTAPGTFILSARSSSLAPDSSF 239

Qy 241 WANHDSKYAMGTSMTATPIVAGNVAQLREHFVKNRGITPKPSLLKAAALIAAGADIGLY 300
Db 240 WANYSKYAYMGTSMTATPIVAGNVAQLREHFVKNRGITPKPSLLKAAALIAAGADIGLY 299

Qy 301 PNGNQGWGRVTLDKSLNVAYNVESLSSTSQKATYSFTATAGKPLKISLWSDAPASTTA 360
Db 300 PSGDQGWGRVTLDKSLNVAYNVEATATGQKATYSFQAQKPLKISLWTDAPGSTTA 359

Qy 361 SVTLVNDLNLVITAPNGTQVGNDFSPYNDNWDGRNNVNFVINAPOSQGTITIEVQAYN 420
Db 360 SVTLVNDLNLVITAPNGQKIVGNDFSPYDNNWDGRNNVNFVINAPOSQGTITIEVQAYN 419

Qy 421 VPVGPQTFSLAIYN 434
Db 420 VPSPQRFSLAIHV 433

RESULT 8

US-09-985-689A-4

; Sequence 4, Application US/09985689A

; Publication No. US20030022351A1

; GENERAL INFORMATION:

; APPLICANT: HATADA, YUJI

; APPLICANT: OGAWA, AKINORI

; APPLICANT: KAGEYAMA, YASUSHI

; APPLICANT: SATO, TSUYOSHI

; APPLICANT: ARAKI, HIROYUKI


```
; APPLICANT: SUMITOMO, NOBUYUKI
; APPLICANT: OKUDA, MITSUYOSHI
; APPLICANT: SAEKI, KATSUHIISA
; TITLE OF INVENTION: Alkaline proteases
; FILE REFERENCE: 215483USO
; CURRENT APPLICATION NUMBER: US/09/985,689A
; PRIOR FILING DATE: 2002-07-01
; PRIOR APPLICATION NUMBER: JP P2000-355166
; PRIOR FILING DATE: 2000-11-22
; PRIOR APPLICATION NUMBER: JP P2001-114048
; PRIOR FILING DATE: 2001-04-12
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 433
; TYPE: PRT
; ORGANISM: Bacillus sp.
US-09-985-689A-4

Query Match      88.2%; Score 1982.5; DB 10; Length 433;
Best Local Similarity 87.18; Pred. No. 1.7e-164;
Matches 378; Conservative 30; Mismatches 25; Indels 1; Gaps 1;

Qy 1 NDVARGIVKADVAQSSYGLYGGQGIIVAVADTGLDGRNDSSMHEAFRGKITALYALGRTN 60
Db 1 NDVARGIVKADVAQNNYGLYGGQGVAVADTGLDGRNDSSMHEAFRGKITALYALGRTN 60
Qy 61 NANDTNGHGHVAGSVLNGSTNKGMAPQANLVFQSIMDSGGGLGGLPSNLQTLFSQAYS 120
Db 61 NASDPNGHGHVAGSVLGN-ALNKGMAPQANLVFQSIMDSGGGLGGLPSNLQTLFSQAWN 119
Qy 121 AGARIHTNSGAAVNGAYTTDSRVDDYVRKNDMTILFAAGNEGPNGGTISAPGTAKNAI 180
Db 120 AGARIHTNSGAPVNGAYTANSRQVDEYVRNDMTVLFAAGNEGPSGTISAPGTAKNAI 179
Qy 181 TVGATENLRPSFGSYADNINHVAQFSSRGPTKGRIPKPDVWAPGTILSARSSLAPDSF 240
Db 180 TVGATENLRPSFGSIADNPNIHQAQFSSRGATGRIPKPDVAPGTFTILSARSSLAPDSF 239
Qy 241 WANHDSKYAYNGGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAAIAGAADIGLGY 300
Db 240 WANYSKYAYNGGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAAIAGAADVGLGY 299
Qy 301 PNGQGWGRVTLKSLNVAVNVSSTLSQKATYSFTATAGKPLKISLWSDAPASTTA 360
Db 300 PNGQGWGRVTLKSLNVAVNVEATATGKATYSFQAQAGKPLKISLWTDAPGSTTA 359
Qy 361 SVTLVNDLNLVITAPNGTQVGNDFTPYNDNWDGRNNVENFINAPQSGTYTIEVQAYN 420
Db 360 SYTLVNDLNLVITAPNGKQVGNDFSPYNDNWDGRNNVENFINAPQSGTYTIEVQAYN 419
Qy 421 VPVGQPTFSLAIVN 434
Db 420 VPSPQPTFSLAIVH 433

RESULT 9
US-10-336-324-10
; Sequence 10, Application US/10336324
; Publication No. US20030176304A1
; GENERAL INFORMATION:
; APPLICANT: Hansen, Peter
; APPLICANT: Bauditz, Peter
; APPLICANT: Mikkelsen, Frank
; APPLICANT: Andersen, Kim
; TITLE OF INVENTION: Protease Variants and Compositions
; FILE REFERENCE: 5349.204-US
; CURRENT APPLICATION NUMBER: US/10/336,324
; CURRENT FILING DATE: 2003-01-03
; PRIOR APPLICATION NUMBER: US/09/512,251A
; PRIOR FILING DATE: 2000-02-24
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 10
; LENGTH: 345
; TYPE: PRT
; ORGANISM: Bacillus
US-10-336-324-10

Query Match      70.4%; Score 1581.5; DB 14; Length 345;
Best Local Similarity 95.3%; Pred. No. 1.2e-129;
Matches 303; Conservative 10; Mismatches 4; Indels 1; Gaps 1;

Qy 1 NDVARGIVKADVAQSSYGLYGGQGIIVAVADTGLDGRNDSSMHEAFRGKITALYALGRTN 60
Db 29 NDVARGIVKADVAQNNFGLYGGQGIIVAVADTGLDGRNDSSMHEAFRGKITALYALGRTN 88
Qy 61 NANDTNGHGHVAGSVLNGSTNKGMAPQANLVFQSIMDSGGGLGGLPSNLQTLFSQAYS 120
Db 89 NANDPENGHGHVAGSVLGN-ATNKGMAPQANLVFQSIMDSGGGLGGLPANLQTLFSQAYS 147
Qy 121 AGARIHTNSGAAVNGAYTTDSRVDDYVRKNDMTILFAAGNEGPNGGTISAPGTAKNAI 180
Db 148 AGARIHTNSGAPVNGAYTTDSRVDDYVRKNDMTILFAAGNEGPSGTISAPGTAKNAI 207
Qy 181 TVGATENLRPSFGSYADNINHVAQFSSRGPTKGRIPKPDVWAPGTFTILSARSSLAPDSF 240
Db 208 TVGATENLRPSFGSYADNINHVAQFSSRGPTDRGRIPKPDVWAPGTFTILSARSSLAPDSF 267
Qy 241 WANHDSKYAYNGGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAAIAGAADIGLGY 300
Db 268 WANHDSKYAYNGGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKAAIAGAADVGLGF 327
Qy 301 PNGQGWGRVTLKSLNV 318
Db 328 PNGQGWGRVTLKSLNV 345

RESULT 10
US-10-403-105-13
; Sequence 13, Application US/10403105
; Publication No. US20030180933A1
; GENERAL INFORMATION:
; APPLICANT: Hansen, Peter K.
; APPLICANT: Bauditz, Peter
; APPLICANT: Mikkelsen, Frank
; TITLE OF INVENTION: Protease Variants and Compositions
; FILE REFERENCE: 5435.200-US
; CURRENT APPLICATION NUMBER: US/10/403,105
; CURRENT FILING DATE: 2003-03-31
; PRIOR APPLICATION NUMBER: US/09/196,281A
; PRIOR FILING DATE: 1998-11-19
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 1332/97
; PRIOR FILING DATE: EARLIER FILING DATE: 1997-11-21
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 13
; LENGTH: 345
; TYPE: PRT
; ORGANISM: Bacillus
US-10-403-105-13

Query Match      70.4%; Score 1581.5; DB 14; Length 345;
Best Local Similarity 95.3%; Pred. No. 1.2e-129;
Matches 303; Conservative 10; Mismatches 4; Indels 1; Gaps 1;

Qy 1 NDVARGIVKADVAQSSYGLYGGQGIIVAVADTGLDGRNDSSMHEAFRGKITALYALGRTN 60
Db 29 NDVARGIVKADVAQNNFGLYGGQGIIVAVADTGLDGRNDSSMHEAFRGKITALYALGRTN 88
Qy 61 NANDTNGHGHVAGSVLNGSTNKGMAPQANLVFQSIMDSGGGLGGLPSNLQTLFSQAYS 120
Db 89 NANDPENGHGHVAGSVLGN-ATNKGMAPQANLVFQSIMDSGGGLGGLPANLQTLFSQAYS 147
Qy 121 AGARIHTNSGAAVNGAYTTDSRVDDYVRKNDMTILFAAGNEGPNGGTISAPGTAKNAI 180
```

Db 148 AGARIHTNSWGPVANGAYTTDSRNVDYVRKNDMTILFAAGNEGPGSGTISAPGTAKNAI 207
QY 181 TVGATENLRPSFGSYADNINHVAQFSSRGPTKDGRIKPDVMAPTGFIILSARSSSLAPDSSF 240
Db 208 TVGATENLRPSFGSYADNINHVAQFSSRGPTKDGRIKPDVMAPTGFIILSARSSSLAPDSSF 267
QY 241 WANHDSKYAYMGSTSMATPIVAGNVAQREHFVKNRGITPSPSLKAAALIAAADIGLY 300
Db 268 WANHDSKYAYMGSTSMATPIVAGNVAQREHFVKNRGITPSPSLKAAALIAAADIGLY 327
QY 301 PNGNCGWGRVTLDKSLNV 318
Db 328 PNGNCGWGRVTLDKSLNV 345

RESULT 11
US-10-090-624-12
; Sequence 12, Application US/10090624
; Publication No. US20020132335A1
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: SHIMOJO, Tomoko
; APPLICANT: ASADA, Kiyozo
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
; FILE REFERENCE: TAKAKURA=6
; CURRENT APPLICATION NUMBER: US/10/090,624
; PRIOR FILING DATE: 2002-03-06
; PRIOR APPLICATION NUMBER: 09/445,472
; PRIOR FILING DATE: 1999-12-06
; PRIOR APPLICATION NUMBER: 151969/1997
; PRIOR FILING DATE: 1997-06-10
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 12
; LENGTH: 659
; TYPE: PRT
; ORGANISM: Thermococcus celer
US-10-090-624-12

Query Match 19.9%; Score 447.5; DB 13; Length 659;
Best Local Similarity 29.8%; Pred. No. 3e-30;
Matches 137; Conservative 68; Mismatches 153; Indels 101; Gaps 18;

QY 8 VKADVAQSSVGLYGQGVAVADTGLDTCGRNDSMHEAPRGKITALLY-ALGRTNNANDTN 66
Db 145 IGADTVNLSGLVDSGVVVAIVDTGIDAN-----HPDLKGVIGYDAVNGRSTPYDDQ 198
QY 67 GHGTHVAGSVLNGSTNK---GNAPQANLVFQSIM--DSGGGLGLPSNLQTLFSAQYSA 121
Db 199 GHGTHVAGSVAGTGSVNSQYIGVAPGAKLVGVKVLGADGSGSVTIIAGVDVWVQNKDKY 258
QY 122 GARI-----HTNSWGAAYNGAYTTDSRNVDYVRKNDMTILFAAGNEGPGSGT 169
Db 259 GIRVNLGLSGQSSDGTDLSQLAVNNAWA-----GIVCVVAGNSGFNYT 306
QY 170 ISAPGTAKNAITVGATENLRPSFGSYADNINHVAQFSSRGPTKDGRIKPDVMAPTGFIILS 229
Db 307 VGSPPAAASKVITVGA-----VDGNDNIAFSRSGPTADGRLKPEVAVPGVDIIA 355
QY 230 ARSSSLAPDSSFWANHDSKYAYMGSTSMATPIVAG-NVAQLREHFVKNRGITPK--PSLLK 286
Db 356 PRAS-----GTSMTGPIINDYITKASGTSMATPHVSGVGLITLQAH-----PSWTPDKVY 404
QY 287 AALIAGA-----ADIGLYPENGQWGRVTLDKSL---NVAYVNESSLSSTSQKATY 335
Db 405 TALIEADIVAPKEIADIAVGA-----GRNVYKAIKYDDYAKLITGSAVDKGSATH 457
QY 336 SFTATAGKLEKSLVWSADAPASTASVTLVNLDNLVITAPNGTQYVGNDFTSFYNDNWDG 395
Db 458 TFDVSGATTVTATLYWD-----TGSSDIDLYLDPNGNE-VDYSYATY----- 500

QY 396 RNNVENVFINAPQSGTYTIEVOAYNVVPGPQTFSLAIVN 434
Db 501 --CFEKVGYNPTAGTWKVVSYK---GAANYQVDVVS 534

RESULT 12
US-10-090-624-1
; Sequence 1, Application US/10090624
; Publication No. US20020132335A1
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: SHIMOJO, Tomoko
; APPLICANT: ASADA, Kiyozo
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
; FILE REFERENCE: TAKAKURA=6
; CURRENT APPLICATION NUMBER: US/10/090,624
; CURRENT FILING DATE: 2002-03-06
; PRIOR FILING DATE: 1999-12-06
; PRIOR APPLICATION NUMBER: 09/445,472
; PRIOR FILING DATE: 1999-12-06
; PRIOR APPLICATION NUMBER: 151969/1997
; PRIOR FILING DATE: 1997-06-10
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 1
; LENGTH: 412
; TYPE: PRT
; ORGANISM: Pyrococcus furiosus
US-10-090-624-1

Query Match 18.2%; Score 409; DB 13; Length 412;
Best Local Similarity 29.8%; Pred. No. 3.5e-27;
Matches 135; Conservative 60; Mismatches 147; Indels 114; Gaps 19;

QY 18 GLYGQGVAVADTGLDTCGRNDSMHEAPRGKITALLYALGRTNNAN-----DTNGHGH 71
Db 22 GYDGSGITIGIITGID-----ASHPDLQGV-----IGWDFVNGRSYPYDDHGHGTH 70
QY 72 VAGSVLNGSTN---KGWAPQANLVFQSIM--DSGGGLGLPSNLQTLFSAQYSAQARI 125
Db 71 VASIAAGTGAASNGKYKGWAPGAKIGIKVLGADGSGSISTIKGVEMAVDNKDKYGIKV 130
QY 126 HTNSWGA-----AVNGAYTTDSRNVDYVRKNDMTILFAAGNEGPGSGTISAP 173
Db 131 INLSLGSQSSDGTDLSQLAVNNAWA-----GLVWVAAGNSGPNKYTIGSP 178
QY 174 GTAKNAITVGATENLRPSFGSYADNINHVAQFSSRGPTKDGRIKPDVMAPTGFIILSARSS 233
Db 179 AAASKVITVGA-----VDKYDVITSSRSGPTADGRLKPEVAVPGNWIITARAS 227
QY 234 LAPDSSFWANHDSKYAYMGSTSMATPIVAGNVAQREHFVKNRGITPK--PSLLKAAALIA 291
Db 228 ---GTSWGPINDYITAAAGTSMATPHVAGIAALLQ-----AHPSTPDKVKTALIE 277
QY 292 GA-----ADIGLYPENGQWGRVTLDKSLNVAIVNESSLSSTSOA-----TYSFT 338
Db 278 TADIVKPEDEIADIAVGA-----GRNAYKAIN--YDNYAKLVFTGYVANKGSGTHQFV 328
QY 339 ATAGKPLKISLVWSDAPASTASVTLVNLDNLVITAPNGTQYVGNDFTSFYNDNWDGRNN 398
Db 329 ISGASFVATLYWDNAN-----SDDLILYDPNGNQ-VDYSYATY-----G 369
QY 399 VENVFINAPQSGTYTIEVOAYNVVPGPQTFSLAIVN 434
Db 370 FEKVGYNPTDGTWTKVVSYS---GSANYQVDVVS 402

RESULT 13
US-10-090-624-4
; Sequence 4, Application US/10090624
; Publication No. US20020132335A1
; GENERAL INFORMATION:

APPLICANT: TAKAKURA, Hikaru
APPLICANT: MORISHITA, Mio
APPLICANT: SHIMOJO, Tomoko
APPLICANT: ASADA, Kiyozo
APPLICANT: KATO, Ikunoshin
TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
FILE REFERENCE: TAKAKURA-6
CURRENT APPLICATION NUMBER: US/10/090,624
CURRENT FILING DATE: 2002-03-06
PRIOR FILING DATE: 1999-12-06
PRIOR APPLICATION NUMBER: 151969/1997
NUMBER OF SEQ ID NOS: 33
SOFTWARE: PatentIn version 3.0
SEQ ID NO 4
LENGTH: 522
TYPE: PRT
ORGANISM: Pyrococcus furiosus
NAME/KEY: misc feature
LOCATION: (428)..(428)
OTHER INFORMATION: Xaa at position 428 is Gly or Val.
US-10-090-624-4

Query Match 18.2%; Score 409; DB 13; Length 522;
Best Local Similarity 29.6%; Pred. No. 4.9e-27;
Matches 135; Conservative 60; Mismatches 147; Indels 114; Gaps 19;
QY 18 GLYGQGIIVAVADTGLDTRNDSSMHEAFRGKITALVALGRTNAN-----DTNGHGTH 71
DB 22 GYDGSGITIGIDGID-----ASHPDLQKV-----IGWVDFVNGRSYFYDDHGTH 70
QY 72 VAGSVLNGSTN-----KGMAPQANLVFQSIW--DSGGGLGLPSNLQTLFSQAYSAGARI 125
DB 71 VASIAAGTGAASNGKYKGMAPGAKLAGIKVLGADGSGSISITIKGVEMAVDNKDKYGIKV 130
QY 126 HTNSGWA-----AVNGAYTTDSRNVDDYVRKNDMTILFAAGNEGPGNGTISAP 173
DB 131 INLSGSSQSDGTDALSQAVNAAWDA-----GLVVVAAGNSGPNKYTIGSP 178
QY 174 GTAKNAITVGATENLRPSFGSYADININHAOFSSRGPTKDGRIKPDVMAFGTILSARSS 233
DB 179 AAASKVITVGA-----VDKYDVITSFSSRGPTADGRLKPEVVAPGNWIIAARAS 227
QY 234 LAPDSSFWANHDSKYAYMGSTMATPIVAGNVAOLREHFVKNRGITPK--PSLLKAALIA 291
DB 228 ---GTSMGQPINDDYTAAPGTSMATPHVAGIAALLQ-----AHPSWTPDKVKTALIE 277
QY 292 GA-----ADIGLGYPNQNGWRVTLKSLNVAIVNESSLSLSOKA-----TYSFT 338
DB 278 TADIVKPEDEIADIAVGA-----GRVNAVYKAIN--YDNVAKLVFTGYVANKGSOHQFV 328
QY 339 ATAGKPLKISLWSDAPASTTASVTLVNDLNLVITAPNGTYVGNDFTSFYNDNWDGRNN 398
DB 329 ISGASFVATLYWDNAN-----SDLDLYDPNGNQ--VDYSYTAYY-----G 369
QY 399 VENVFIPAPQSGTYTIEVQAYNVPGQTFSLAIVN 434
DB 370 FEKVGYNPTDGTWIKVSVS---GSANYQVDVVS 402

RESULT 14
US-10-090-624-16
Sequence 16, Application US/10090624
Publication No. US20020132335A1
GENERAL INFORMATION:
APPLICANT: TAKAKURA, Hikaru
APPLICANT: MORISHITA, Mio
APPLICANT: SHIMOJO, Tomoko
APPLICANT: ASADA, Kiyozo
APPLICANT: KATO, Ikunoshin
TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE

FILE REFERENCE: TAKAKURA-6
CURRENT APPLICATION NUMBER: US/10/090,624
CURRENT FILING DATE: 2002-03-06
PRIOR APPLICATION NUMBER: 09/445,472
PRIOR FILING DATE: 1999-12-06
PRIOR APPLICATION NUMBER: 151969/1997
NUMBER OF SEQ ID NOS: 33
SOFTWARE: PatentIn version 3.0
SEQ ID NO 16
LENGTH: 654
TYPE: PRT
ORGANISM: Pyrococcus furiosus
US-10-090-624-16
Query Match 18.2%; Score 409; DB 13; Length 654;
Best Local Similarity 29.6%; Pred. No. 6.9e-27;
Matches 135; Conservative 60; Mismatches 147; Indels 114; Gaps 19;
QY 18 GLYGQGIIVAVADTGLDTRNDSSMHEAFRGKITALVALGRTNAN-----DTNGHGTH 71
DB 154 GYDGSGITIGIDGID-----ASHPDLQKV-----IGWVDFVNGRSYFYDDHGTH 202
QY 72 VAGSVLNGSTN-----KGMAPQANLVFQSIW--DSGGGLGLPSNLQTLFSQAYSAGARI 125
DB 203 VASIAAGTGAASNGKYKGMAPGAKLAGIKVLGADGSGSISITIKGVEMAVDNKDKYGIKV 262
QY 126 HTNSGWA-----AVNGAYTTDSRNVDDYVRKNDMTILFAAGNEGPGNGTISAP 173
DB 263 INLSGSSQSDGTDALSQAVNAAWDA-----GLVVVAAGNSGPNKYTIGSP 310
QY 174 GTAKNAITVGATENLRPSFGSYADININHAOFSSRGPTKDGRIKPDVMAFGTILSARSS 233
DB 311 AAASKVITVGA-----VDKYDVITSFSSRGPTADGRLKPEVVAPGNWIIAARAS 359
QY 234 LAPDSSFWANHDSKYAYMGSTMATPIVAGNVAOLREHFVKNRGITPK--PSLLKAALIA 291
DB 360 ---GTSMGQPINDDYTAAPGTSMATPHVAGIAALLQ-----AHPSWTPDKVKTALIE 409
QY 292 GA-----ADIGLGYPNQNGWRVTLKSLNVAIVNESSLSLSOKA-----TYSFT 338
DB 410 TADIVKPEDEIADIAVGA-----GRVNAVYKAIN--YDNVAKLVFTGYVANKGSOHQFV 460
QY 339 ATAGKPLKISLWSDAPASTTASVTLVNDLNLVITAPNGTYVGNDFTSFYNDNWDGRNN 398
DB 461 ISGASFVATLYWDNAN-----SDLDLYDPNGNQ--VDYSYTAYY-----G 501
QY 399 VENVFIPAPQSGTYTIEVQAYNVPGQTFSLAIVN 434
DB 502 FEKVGYNPTDGTWIKVSVS---GSANYQVDVVS 534

RESULT 15
US-10-112-488-39
Sequence 39, Application US/10112488
Publication No. US20030082746A1
GENERAL INFORMATION:
APPLICANT: KIKUCHI, Yoshiaki
APPLICANT: DATE, Masayo
APPLICANT: UMEZAWA, Yukiko
APPLICANT: YOKOYAMA, Keiichi
APPLICANT: MATSUI, Hiroshi
TITLE OF INVENTION: PROCESS FOR PRODUCING TRANSGLUAMINASE
FILE REFERENCE: 219286USOCONT
CURRENT APPLICATION NUMBER: US/10/112,488
CURRENT FILING DATE: 2002-04-01
PRIOR APPLICATION NUMBER: PCT/JP00/06780
PRIOR FILING DATE: 2000-09-29
PRIOR APPLICATION NUMBER: JP2000-280098
PRIOR FILING DATE: 2000-06-28
PRIOR APPLICATION NUMBER: JP11-280098
PRIOR FILING DATE: 1999-09-30
NUMBER OF SEQ ID NOS: 70

```
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 39
; LENGTH: 1079
; TYPE: PRT
; ORGANISM: Streptomyces albobogiseolus
US-10-112-488-39

Query Match      16.3%; Score 366; DB 14; Length 1079;
Best Local Similarity 31.8%; Pred. No. 8.1e-23;
Matches 148; Conservative 55; Mismatches 175; Indels 90; Gaps 20;

QY   3 VAR----GIVKADVAQS-----SYCLYGQGGQIVAVADTGLDTRNDSSMHEAFRG 48
Db   160 VARVWLDGVRKASLDTSVGQIGTPKAWAGYDGKGVKIAVLDTGVD-----ATHPDLKG 213
QY   49 KITALYALGRTNNANTNGHGHVAGSVLNGS-----TNKGAPQANLVFQSIMDSGGGL 104
Db   214 QVTASKNFTSAPTGTGDVWGHGTHVASIAAGTGAQSGTYKGVAPGAKILNGKVLDDAG-- 271
QY   105 GGLPSNLQTLFSQAYSAGARIHTNSWGAAVNGAYTTDSRNVDYVRK--NDMTILF--AA 160
Db   272 FGDDSGILAGMEWAQAQGADIWNMSLG-----GMDTPETDPLEAANDKLSAEKGILFAIAA 327
QY   161 GNEGPNGGTISAPGTAKNAITVGATENLRPSFGSYADNINHVAQFSSRGP-TKDGRIKPD 219
Db   328 GNEGPO--SIGSPGSADSALTVGA-----VDDKDKLADFSSTGPRLGDGAVKRPD 374
QY   220 VMAPGTFILSARSLAPDSFWANHDSKYAYMGGTSMATPIVAGNVAQLREHFVKNRGIT 279
Db   375 LTAPGVDTIAASAKGNNDIAKEVGEKPGAYWTISGTSNATPHVAGAAALLKQOHPE----- 429
QY   280 PKPSLLKAALIAGAADIGLG-YPNGNGWGRVTLDKSLNVAAYYNSSSLSS----- 328
Db   430 WKYAEKLGALTASTKD---GKYTPFEQSGSRVQVDKAITQTVIABPVSLSFGVQWPHAD 486
QY   329 ---TSOKATYSFTATAGKPLKISLVMSD-----APAS--TTASVTIANDNLNLTAP-NG 377
Db   487 DKPVTKLTYRNLTGTEEDVTLKLTSTATGPKGAAPAGFTLGCSTL-----TVPANG 538
QY   378 TQYVGNDFTPSYNDNDWNGRNNVENFINAPQS---GTYTIEVQAYNV 421
Db   539 TASVDVTADTRLGGAVDGTYSAYVATGAGQSVRTAAAVEREVESINV 586
```

Search completed: March 10, 2004, 15:00:19
Job time : 41 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2004, 14:51:09 ; Search time 20 Seconds
(without alignments)
2087.356 Million cell updates/sec

Title: US-09-985-689a-1-COPY
Perfect score: 2247
Sequence: 1 NDVARGIVKADVAQSYGLY.....EVQAVNVPGVQTFSLAIVN 434

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR.78.*

- 1: PIR1.*
- 2: PIR2.*
- 3: PIR3.*
- 4: PIR4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|-------|-------------|--------|----------|---------------------|
| 1 | 519.5 | 23.1 | 1743 | 2 T18279 | multidrug resistanc |
| 2 | 492 | 21.9 | 1905 | 2 T18267 | multidrug resistanc |
| 3 | 347.5 | 15.5 | 444 | 2 B83891 | intracellular alka |
| 4 | 320.5 | 14.3 | 442 | 2 A69887 | intracellular alka |
| 5 | 311.5 | 13.9 | 806 | 2 A41341 | microbial serine p |
| 6 | 303.5 | 13.5 | 1398 | 2 T28159 | pyrolysin (EC 3.4 |
| 7 | 281 | 12.5 | 419 | 1 S25835 | subtilisin-type pr |
| 8 | 280 | 12.5 | 799 | 2 G83753 | subtilisin (EC 3.4 |
| 9 | 279.5 | 12.4 | 1345 | 2 T29090 | surface layer-asso |
| 10 | 278 | 12.4 | 580 | 2 S11890 | serine proteinase |
| 11 | 277 | 12.3 | 420 | 1 S23407 | subtilisin (EC 3.4 |
| 12 | 265 | 11.8 | 715 | 2 JC4908 | alkaline serine pr |
| 13 | 264 | 11.7 | 1331 | 2 A72647 | probable surface l |
| 14 | 262.5 | 11.7 | 513 | 1 A35742 | aqualysin (EC 3.4 |
| 15 | 256.5 | 11.4 | 757 | 2 C84120 | subtilisin-type pr |
| 16 | 253.5 | 11.3 | 894 | 2 F69730 | cell wall-associat |
| 17 | 251 | 11.2 | 627 | 2 D75393 | serine proteinase |
| 18 | 246.5 | 11.0 | 402 | 1 J70332 | alkaline proteinase |
| 19 | 244.5 | 10.9 | 534 | 1 U80173 | alkaline proteinase |
| 20 | 242.5 | 10.8 | 401 | 2 I39974 | serine proteinase |
| 21 | 240 | 10.7 | 380 | 2 A49778 | high-alkaline seri |
| 22 | 240 | 10.7 | 488 | 2 A11930 | proteinase (import |
| 23 | 238 | 10.6 | 519 | 2 S71451 | halolysin R4 (EC 3 |
| 24 | 235.5 | 10.5 | 382 | 1 SUBSN | subtilisin (EC 3.4 |
| 25 | 235 | 10.5 | 378 | 2 A33973 | high-alkaline seri |
| 26 | 235 | 10.5 | 382 | 2 I39780 | subtilisin (EC 3.4 |
| 27 | 234 | 10.4 | 910 | 2 C69456 | subtilisin sendai |
| 28 | 233 | 10.4 | 1374 | 2 D72593 | hypothetical prote |
| 29 | 231 | 10.3 | 321 | 1 S27501 | alkaline proteinase |

| | | | | | |
|----|-------|------|------|----------|---------------------|
| 30 | 230.5 | 10.3 | 379 | 1 SUBSCL | subtilisin (EC 3.4 |
| 31 | 230.5 | 10.3 | 525 | 2 G84406 | halolysin (importe |
| 32 | 230.5 | 10.3 | 601 | 2 JC4576 | serine proteinase |
| 33 | 226.5 | 10.1 | 1052 | 2 T17093 | intraluminal subti |
| 34 | 226.5 | 10.1 | 1118 | 2 H97298 | subtilisin like pr |
| 35 | 225.5 | 10.0 | 381 | 2 JH0778 | subtilisin (EC 3.4 |
| 36 | 225.5 | 10.0 | 613 | 2 S75976 | hypothetical prote |
| 37 | 225.5 | 10.0 | 1167 | 1 A35066 | streptococcal Csa |
| 38 | 224.5 | 10.0 | 769 | 2 D86335 | T20H2.6 protein - |
| 39 | 223.5 | 9.9 | 381 | 1 SUBSS | subtilisin (EC 3.4 |
| 40 | 223.5 | 9.9 | 381 | 1 SUBSI | subtilisin (EC 3.4 |
| 41 | 223.5 | 9.9 | 381 | 2 JQ1487 | subtilisin (EC 3.4 |
| 42 | 222.5 | 9.9 | 275 | 2 JC1085 | subtilisin (EC 3.4 |
| 43 | 222 | 9.9 | 384 | 2 JC4802 | alkaline proteinase |
| 44 | 221.5 | 9.9 | 272 | 2 A23624 | subtilisin (EC 3.4 |
| 45 | 221 | 9.8 | 1036 | 2 JC5568 | serine proteinase |

ALIGNMENTS

RESULT 1
T18279
Multidrug resistance transport protein - slime mold (Dictyostelium discoideum)
C:Species: Dictyostelium discoideum
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T18279
R:Shaulsky, G.; Loomis, W.F.
Submitted to the EMBL Data Library, June 1996
A:Reference number: Z18855
A:Accession: T18279
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: mRNA
A:Residues: 1-1743 <SHA>
A:Cross-references: EMBL:U60086; NID:g1399914; PID:g1399915; PIDN:AA03331.1
C:Genetics:
A:Gene: tagC

Query Match 23.1%; Score 519.5; DB 2; Length 1743;
Best Local Similarity 28.4%; Pred. No. 3.8e-25;
Matches 166; Conservative 78; Mismatches 158; Indels 183; Gaps 24;

| | | | |
|----|-----|--|-----|
| QY | 19 | LYGQOIIVADTGLDTCR---NDS-----SMHEAFRGKITALYALGRNNDANDNGH | 68 |
| DB | 314 | LRGKGQILSIADTGLDGSCHCFSDSKYPIPLNSVNLNHR-KVVTYITTSDDSKVDGH | 372 |
| QY | 69 | GTHVAGSVLG-----NGSTNKGMAPQANLVFQSDGGGLGL--PSNLQTLFSGAY | 119 |
| DB | 373 | GTHICSGAAGTPEDSSVNISSFGSLATDAKAF---FDLASGSSSLTPPSDLKQLYQPLY | 429 |
| QY | 120 | SAGARIHTNSGA-----AVNGAYTDSNRVDYVRKN-DMTILPAAGNEGNGGTIS--A | 172 |
| DB | 430 | DAGARVHCDSWGVSVEGVTGYSYSDTASIDDFLTFHPDFILRAAGN---NEOYLSLLT | 486 |
| QY | 173 | PCTAKNAITVGATENLR-----PSFGSYADNT-----PSFGSYADNT----- | 199 |
| DB | 487 | QSTAKNVITVGAHQTHENYLTGPNYINYOSSVDINQELICDFDSRYCNYTTAQCLLES | 546 |
| QY | 200 | -----NHVAQFSSRGPTKDGRIKPDVMAPGFTIL | 228 |
| DB | 547 | NATTGLASCCPTLLRKSVIDAANTQPLLNYENNICSFSKGPETHDGRMKPALVAPGEYIT | 606 |
| QY | 229 | SARSSLA-----PDSSFWAHDSDKYVMGGTGVATPIVAGNVAQLREH----- | 272 |
| DB | 607 | SARSGANTTQCCGGSL-PMTNALLA-IGSTWATSPAAATILRLQYLVDGYPTGSI | 664 |
| QY | 273 | VNVRGITPKPSLLKAALIAGA-----ADIGLYPNGN-----QGWGRVT | 311 |
| DB | 665 | VESNKLPQTGSLIKALMINNAQLLNGTLPQITSSSITYPSNQVFENFAGSLVQGWGAI | 724 |
| QY | 312 | LDKSLNAVYVNESS-----SLSQKATYSFT-- | 338 |
| DB | 725 | MSNWLHLVNNNNNNNNKTSIDGITKFDGIGGLDLRLVKPNQWKBESUSTGNTSYCTFYK | 784 |

R; Voorhorst, W.G.B.; Eagen, R.I.L.; Geerling, A.C.M.; Platteeuw, C.; Siezen, R.J.; Vos, J. Biol. Chem. 271, 20426-20431, 1996
A; Title: Isolation and characterization of the hyperthermostable serine protease, pyrolysin
A; Reference number: Z20481; MUID:196355370; PMID:8702780
A; Accession: T28159
A; Status: preliminary; translated from GB/EMBL/DBJ
A; Molecule type: DNA
A; Residues: 1-1398 <VOO>
A; Cross-references: EMBL:U55835; NID:gl556462; PID:gl556463; PIDN:AAB09761.1
A; Experimental source: DSM3638
C; Geneticks:
A; Gene: pls
C; Keywords: hydrolase; serine proteinase

Query Match 13.5%; Score 303.5; DB 2; Length 1398;
Best Local Similarity 26.6%; Pred. No. 1.9e-11;
Matches 139; Conservative 57; Mismatches 178; Indels 149; Gaps 22;

Qy 21 GQGIIVAVADTGLDGTGRNDS-----SMHEAFRGKITALYALGRNNAN----- 63
Db 301 GNGYDIAYVDTLDYDFTDEVLPGQNVTVYDVAVFYYGGLNYYLAEIDPNGEYAVFGW 360

Qy 64 DTNGHGHVAGSVLNGSTN-----KGMAPQAN 91
Db 361 DGHGHGHVAGTVAGVDSNNDWDLMSYSGEWFSLRYGHYTNVTTDVTQGVAPQAQ 420

Qy 92 LVFQSIMDSGGGLGLPSNLQTLFQAVSAGARIHTNSWGAAVNGAYT--TDSRN--YDD 147
Db 421 IMAIRVLR--DGRGSMWDIEGM--TYAATHGADVISMISLGG--NAPYLDGTDPSVAVDE 476

Qy 148 VYRKNDWTILFAGNEGPNNGGTISAPGTAKNAITVGTENLRPSGVSAD----- 197
Db 477 LTBKYGVVFIAGNEGPNINGVSGVATKAITVGA--VPINVGTVVQALGVDPDYVG 535

Qy 198 -----NINHAQFSSRGPTKDGRIKPDVNPAGTFFILSRSLAPDSSFWANHDSKYA 249
Db 536 FYVFPAYTV--RTAFSSRGPRIDGIEKPNVAVGVIGYSSLPWIMGGADF----- 585

Qy 250 YNGGTSMATPIVAGNVAQLREHFVKNRGITPKPSLLKALIAAGADI-----GLGYPNG 303
Db 586 -MSGTSMATPHVSGVALLISG--PKPEGIYYPDIKKVLESAGATWLEGDPYTGOKYTEL 643

Qy 304 NQGWGRTVTDKSLNVAVNESLSLTSQKATSFYATAGKPLKISLWSDAPASTA--- 360
Db 644 DQGHGLVNVTKSWEI-----LKAINGTTTLPIVDHWADKSYSDFAEYL 685

Qy 361 SVTLVNDLNLVITAPN-----GTOYGN-----DFTSPYNDW-----DG-----RNVENVF 403
Db 686 GVDVIRGLYARNIPDIVEWHIKYVGDTEYRTFEIYATEPWIKPFVSGVILENNTEFVL 745

Qy 404 -----INAPQSGTY-----TIEVQAVNPVPGPQTF 429
Db 746 RVKYDVEGLEPGLVGRILIDPTTPVIEDEILNTIVIEKET 788

RESULT 7
S25835
subtilisin (EC 3.4.21.62) precursor - Bacillus sp. (strain TA41)
C; Species: Bacillus sp.
C; Date: 22-Nov-1993 #sequence_revision 20-Feb-1995 #text_change 18-Jun-1999
C; Accession: S25835
R; Davail, S.; Feller, G.; Narinx, E.; Gerday, C.
Gene 119, 143-144, 1992
A; Title: Sequence of the subtilisin-encoding gene from an antarctic psychrotroph Bacillus
A; Reference number: S25835; MUID:93012966; PMID:1398082
A; Accession: S25835
A; Molecule type: DNA
A; Residues: 1-419 <DAV>
A; Cross-references: EMBL:X63533; NID:g40198; PIDN:CAA45096.1; PID:g40199
C; Superfamily: subtilisin; subtilisin homology
C; Keywords: extracellular protein; hydrolase; serine proteinase
F; 1-23/Domain: signal sequence #status predicted <SIG>
F; 24-110/Domain: propeptide #status predicted <PRO>

F; 111-419/Product: microbial serine proteinase #status predicted <MAT>
F; 135-373/Domain: subtilisin homology <SBT>
F; 144,184,359/Active site: Asp, His, Ser #status predicted

Query Match 12.5%; Score 281; DB 1; Length 419;
Best Local Similarity 33.0%; Pred. No. 1e-10;
Matches 87; Conservative 38; Mismatches 97; Indels 42; Gaps 13;

Qy 21 GQGIIVAVADTGLDGTGRNDSME--EAFRGKITALYALGRT-----FVVGTFDNTDSCDRQGHGTHVAGSA 189
Db 135 GAGINTAVLDTGVNTNHPDLSNNVEQCK-----FVVGTFDNTDSCDRQGHGTHVAGSA 189

Qy 77 LKNGSTNK-----GMAPOANL-VFOSIMDSGGGLG-GLPSNLQTLFQAVSAGARIHTN--SW 130
Db 190 LANGGTGSGVYVAPADLWAYKVLGDGSGYADDAEAIHAGDOATALNTKVINMGL 249

Qy 131 GAAVNGAYTTDSNVDDYVRKNDMTILFAAGNEGPNNGGTISAPGTAKNAITVGTENLRP 190
Db 250 GSGGESLIT---NAVYAYDKGVLLIAAGNSGPKPGSIGYPGALVNAVVAALNTIQ 306

Qy 191 SFGSYADNINHVAQFSRGPTKDG-----RIKPDVNPAGTFFILSRSLAPDSSFWANH 244
Db 307 N-GTY-----RVADFSRGHKRTAGDYVIQGDVEISAPGAAYST-----W--F 348

Qy 245 DSKYAYMGTSMATPIVAGNVAQL 268
Db 349 DGGYATISGTSMAFRAAGLAAKI 372

RESULT 8
G83753
subtilisin-type proteinase (EC 3.4.21.-) vpr precursor [similarity] - Bacillus halodura
C; Species: Bacillus halodurans
C; Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
C; Accession: G83753
R; Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fuji, F.; Hir
Nucleic Acids Res. 28, 4317-4331, 2000
A; Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
A; Reference number: A83650; MUID:20512582; PMID:11058132
A; Accession: G83753
A; Status: preliminary
A; Molecule type: DNA
A; Residues: 1-799 <STO>
A; Cross-references: GB:AP001510; GB:BA000004; NID:g10173440; PIDN:BA04550.1; GSPDB:GNO
A; Experimental source: strain C-125
C; Geneticks:
A; Gene: vpr
C; Superfamily: microbial serine proteinase vpr; subtilisin homology
C; Keywords: hydrolase; serine proteinase
F; 1-29/Domain: signal sequence #status predicted <SIG>

Query Match 12.5%; Score 280; DB 2; Length 799;
Best Local Similarity 24.7%; Pred. No. 2.8e-10;
Matches 117; Conservative 50; Mismatches 135; Indels 172; Gaps 16;

Qy 18 GLYQGGIIVAVADTGLDGTGRNDSMEAFRGKITALYALGRNNANDT-----NGH 68
Db 171 GYTGEGITVAITDGTVDYTHFD--LVHAF--GDYKGMDFDINDDPQETPPGDPGRGIETH 227

Qy 69 GTHVAGSVLNGSTNKGMAPQANLVFOSIMDSGGGLGLPSNLQTLFQAVSAGARIHTN 128
Db 228 GTHVAGVVAAGLI--KGVAPDANLLAYRVLGPGG--RGSTAGVIAGIERAVQGDADIMNL 284

Qy 129 SWGAANVAGYTTDSNRVDDYVRKNDMTILFAAGNEGPNNGGTISAPGTAKNAITVGTENL 188
Db 285 SLGNTLNDPDPFATSIAL--DWMAEGVVAVTNSNGSPNNWTVGSPGTSRDAISVGAT--- 340

Qy 189 RPSFGSY-----ADNLNH----- 201
Db 341 RLPTNPKYKASVFTSDGIDYPSADIMFPSPDEELLEDGETEYFAFAGLXPGDFEGVDVE 400

Qy 202 ----- 201

Db 401 GKIALIVRGEIPFVEKAENAKAGAVGAIYNNVAGVQVTPCLAIPTIMLSNEDGLKMR 460
QY 202 -----VAQFSRGPT-KDGRKPDVMAFGTIFLSARSGLAPD 237
Db 461 NELENGQNTVFSIEBPKLVGETVADFSGRPVMTWMKPDVSPAGVAIVSTIPHOQD 520
QY 238 SSWANHDSKIAYMGTSMTPTVAGNVAQLREHFVKNGRITPKPSLLKAAIAGAADI- 296
Db 521 DPT-----GYGSRQGTSMASPHVGAALLLEAH-PNWGV---DHVKAALMNTAENLV 569
QY 297 ---GLGYPNGNQGRVTLDKSLNVAYNVNESLSTSQATY-SPTATACKPLK 346
Db 570 DENGNYPHNTQAG-----SIRIVDAIESETLVTFGSHSGFTFKERKQOVE 617

RESULT 9
T29090
surface layer-associated STABLE proteinase - Staphylothermus marinus
N/Alternate names: hyperthermostable proteinase
C/Species: Staphylothermus marinus
C/Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 02-Sep-2000
C/Accession: T29090
R/Mayr, J.; Lupas, A.; Kellermann, J.; Eckerskorn, C.; Baumeister, W.; Peters, J.
Curr. Biol. 6, 739-749, 1996
A/Title: A hyperthermostable protease of the subtilisin family bound to the surface layer
A/Reference number: Z20559; MUID:96385442; PMID:8793300
A/Accession: T29090
A/Status: preliminary; translated from GB/EMBL/DDBJ
A/Molecule type: DNA
A/Residues: 1-1345 <MAY>
A/Cross-references: EMBL:U57968; NID:G1374755; PID:G1374756; PIDN:AAB02323.1
A/Experimental source: strain F1
C/Function:
A/Description: probably serves an exodigestive function related to the organism's energy
A/Note: stoichiometric S-layer component

Query Match 12.4%; Score 279.5; DB 2; Length 1345;
Best Local Similarity 29.9%; Pred. No. 6.2e-10;
Matches 99; Conservative 45; Mismatches 106; Indels 81; Gaps 14;
QY 46 FRKITALYALGRNNANDTNGHGHVA-----GSVL-----GNGSTNK-GMAPQANLV 93
Db 445 YQGRYLAL-----VSDPHGHGTSVATVIAISGRVLYDYGDKLYRIMGVAPGAKI- 495
QY 94 FQSIMDSGGGLGIPSNLQTLFQVAGSAG-----ARHTNSM 130
Db 496 -----AGDALLGNILV--EAWLAGNIVTEEDGVVYLSLDPFGPHRADIIISNW 546
QY 131 GAAVNGAYTTDSRNV-----DYVKNDMTILFAAGNEPNGGTISAPGTAX 177
Db 547 GSIYINFLQFPQIDYRSSFMDDEILAIRNYLIGDHVTIVFAAGNEGPGYSSNGAPGTGL 606
QY 178 NAITVGATE--NLRPSFG---SYADNINHVAQPSRGPDKGRIPKPDVMAQGTIFLSARS 232
Db 607 LVITAGASTLWDYTRIYGYEGVAD---EVIPSSRGPTQGGYKPKPDIVNIGAFEWASTR 663
QY 233 SLAPDSSFWANHDSKYAYMGTSMTPTVAGNVAQLREHFVKNGRITPKPSLLKAAIAGA 292
Db 664 TI-DGRGYGAQPD---VFGGTSEATPYTSGTLALVQAYKEYVNTTDPDVTAKILKSS 718
QY 293 AADIGLGPNGNQGRVTLDKSLNVAYNVNE 323
Db 719 AKDI--WYPAFSGSGRVDALKAADTVFISE 747

RESULT 10
S11890
serine proteinase (EC 3.4.21.-) precursor, extracellular - Xanthomonas campestris pv. ca
N/Alternate names: subtilisin-related proteinase
C/Species: Xanthomonas campestris pv. campestris
C/Date: 21-Nov-1993 #sequence_revision 07-Feb-1997 #text_change 03-Dec-1999
C/Accession: S11890
R/Liu, Y.N.; Tang, J.L.; Clarke, B.R.; Dow, J.M.; Daniels, M.J.

Mol. Gen. Genet. 220, 433-440, 1990
A/Title: A multipurpose broad host range cloning vector and its use to characterise an e
A/Reference number: S11890; MUID:90251253; PMID:2187155
A/Accession: S11890
A/Molecule type: DNA
A/Residues: 1-580 <LIU>
A/Cross-references: EMBL:X51635; NID:948533; PIDN:CAA35962.1; PID:948534
A/Experimental source: Xanthomonas campestris pv. campestris
A/Note: the sequence from Fig. 4 is inconsistent with that from Fig. 3 in having 205-Ala
C/Superfamily: subtilisin; subtilisin homology
C/Keywords: extracellular protein; hydrolase; serine proteinase
F1-32/Domain: signal sequence #status predicted <Sig>
F168-423/Domain: subtilisin homology <SRT>

Query Match 12.4%; Score 278; DB 2; Length 580;
Best Local Similarity 26.0%; Pred. No. 2.5e-10;
Matches 126; Conservative 56; Mismatches 147; Indels 156; Gaps 24;
QY 21 GQGIIVAVDTGL-----DTGRNDSMHEAFRGKITALYALGRNNAND----- 64
Db 168 GSGTVAVIDTGITSHADLANILAGYDFISDATARDGNGRDSNAADGDMVAANECCA 227
QY 65 -----TNGHGHVAGSVLGNSTNGKMAPOA-----NLVFAQSIMD 99
Db 228 GIPAASSWHGTHVAGIVAAVTNTTGVAGTAYGAKVVPVVLGKCGSLSDIADAIWA 287
QY 100 SGGGLGGLPSNLQ--TLFSQAYSAGARIHNSWGAANGAYTTDSRNVDYVRKNDMTIL 157
Db 288 SGGTVSGIPANANPAEIVNMSLGGGSCSTTMQN-AINGAVSRGT-----TVV 334
QY 158 FAAGNEPNGGTISAPGTAKNAITVGATEN--LRPFSGSYADNINHVAQPSRGPRTKQGR 215
Db 335 VAANDASNVSG-SLPANCANVIAVAATTSAGAKASINFTGTI----- 377
QY 216 IKPDVMAQGTIFLSARS--LAPDSSFWANHDSKYAYMGTSMTPTVAGNVAQLRHFV 273
Db 378 ---DVSAPGSSILSTLNSGTTTSGS-----ASYASYNGTSMASPHVAGVVALVQS--V 425
QY 274 KNRGITPK--PSLTK--AALIAGRADIGLGVPGNQMGWRVTLDKSLNVAYNVNES-- 325
Db 426 APTALTPAAVETLLKNTARALPGAC-----SGCGAGIVNADAAYTAA-INGSGGGG 477
QY 326 -----SLSTSQKATYSTATAGKPLKLSLVMSDAPASTTASVTL---VND 367
Db 478 GGGNTLNGTPTVTLGGAATGAELNYTITVPAG-----SGTLTVTTSGGSGD 523
QY 368 LNLVI---TAPNGTQYVGNFTSPYNDNWDGRNVENVFINAQSGTYTTEVQAYNVPVG 424
Db 524 ADLYVRAGSAPTDSAYT-----CRFYRS-----GNAETCTITAP-SGYIYVRLKAYS----- 569
QY 425 PQTFES 429
Db 570 --TFS 572

RESULT 11
S23407
subtilisin (EC 3.4.21.62) 1 precursor - Bacillus sp. (strain TA39)
C/Species: Bacillus sp.
C/Date: 04-Dec-1992 #sequence_revision 04-Dec-1992 #text_change 18-Jun-1999
C/Accession: S23407
R/Narinx, E.; Davail, S.; Feller, G.; Gerday, C.
Biochim. Biophys. Acta 1131, 111-113, 1992
A/Title: Nucleotide and derived amino acid sequence of the subtilisin from the antarctic
A/Reference number: S23407; MUID:92256481; PMID:1581352
A/Accession: S23407
A/Molecule type: DNA
A/Residues: 1-420 <NAP>
A/Cross-references: EMBL:X62369; NID:940200; PIDN:CAA44227.1; PID:940201
C/Genetics:
A/Gene: sub1
C/Superfamily: subtilisin; subtilisin homology
C/Keywords: extracellular protein; hydrolase; serine proteinase

F:1-23/Domain: signal sequence #status predicted <SIG>
 F:24-111/Domain: propeptide #status predicted <PRO>
 F:112-420/Product: microbial serine proteinase #status predicted <MAT>
 F:136-374/Domain: subtilisin homology <SBT>
 F:145,185,360/Active site: Asp, His, Ser #status predicted

Query Match 12.3%; Score 277; DB 1; Length 420;
 Best Local Similarity 31.2%; Pred. No. 1.9e-10;
 Matches 98; Conservative 44; Mismatches 120; Indels 52; Gaps 16;

QY 21 GGGQIVAVADTGLDGRNDSSMHEAFRGKITAL--YALGRT---NNANDTGHGTHVAGS 75
 DB 136 GGGINIAVLDGTGNTN-----HPDLNNVEOCKFTVGTITNNSCCTDRQGHGTHVAGS 189
 QY 76 VLNGSTNK---GMAPOANL-VFQSIMDSGGGLG-GLFSLNQLTFLPSQAYSAGARIHTN-S 129
 DB 190 ALADGGTGNVGVAPADADLWAYKVLGDDSGYADDIAAIRHAGDQATALNTKVINNS 249
 QY 130 WGAANVGATTSRNVDDYVRKNDMTILPAAAGNEGNGGTISAPGTAKNAITVGATENLR 189
 DB 250 LGSSESSITNAVN---YSYKNGVLIIIAAGNSGPGYQSGIGYPGALVNAVAVALEN-K 305
 QY 190 PFGSGYADNINHVAQFSSRGPT-KDG-----RIKPDVMAPGTFILSARSSLAPDSSFVAN 243
 DB 306 VENGTY-----RVADFSSRGYSWTGDYAIQKGDVEISAPGAIIYST-----N-- 348
 QY 244 HDSKIAYMGSTWATPIVAGNVAQLREHFVKRGITPKPSLLKALIAAGADIGLGPNG 303
 DB 349 FDGGYATISGTSNASPHAAGLAAKTIWAQYSPASNVDRGELQYRAY---ENDILSGYYAG 405
 QY 304 -----NQGGGRVTL 312
 DB 406 YGDDFASGFGFATV 419

RESULT 12
 JC4908
 alkaline serine proteinase (EC 3.4.-.-) I precursor - Alteromonas sp.
 A:Alternate names: subtilase
 C:Species: Alteromonas sp.
 C:Date: 01-Nov-1996 #sequence_revision 01-Nov-1996 #text_change 08-Oct-1999
 C:Accession: JC4908
 F:tsujiibo, H.; Miyamoto, K.; Tanaka, K.; Kaidzu, Y.; Imada, C.; Okami, Y.; Inamori, Y.
 Biosci. Biotechnol. Biochem. 60, 1284-1288, 1996
 A:Title: Cloning and sequence analysis of a protease-encoding gene from the marine bacterium *Alteromonas*
 A:Reference number: JC4908; MUID:97141200; PMID:8987544
 A:Accession: JC4908
 A:Status: Preliminary
 A:Molecule type: DNA
 A:Cross-references: DDBJ:D38600; NID:G1536787; PIDN:BAAL8912.1; PID:d1019647; PID:g21602
 A:Experimental source: strain O-7
 C:Comment: This enzyme belongs to class I subtilisin-like family. It is a chelator-sensitized subtilisin.
 C:Genetics:
 A:Gene: aprI
 C:Superfamily: subtilisin homology
 C:Keywords: hydrolase
 F:1-40/Domain: signal sequence #status predicted <SIG>
 F:41-150/Domain: amino-terminal propeptide #status predicted <ATP>
 F:151-496/Product: alkaline serine protease
 F:182-452/Domain: subtilisin homology <SBT>
 F:497-715/Domain: carboxyl-terminal propeptide #status predicted <CTP>
 F:739-294,335-372,478-481/Disulfide bonds: #status predicted

Query Match 11.8%; Score 265; DB 2; Length 715;
 Best Local Similarity 25.8%; Pred. No. 2.2e-09;
 Matches 124; Conservative 48; Mismatches 170; Indels 138; Gaps 22;

QY 21 GGGQIVAVADTGLDGRNDSSMHEAFRGKITALYALGRTNAND-----TNG---- 67
 DB 182 GGGVAVAVLDGTGRPHLDLDANILPGYDMISNTFVANDGGARDNDARDPQDVAITRECGT 241
 QY 68 -----HGTHVAG-----SVLNGSGTNKGMAPQANLVFQSIMDSGGGLGLP 108

F:1-23/Domain: signal sequence #status predicted <SIG>
 F:24-111/Domain: propeptide #status predicted <PRO>
 F:112-420/Product: microbial serine proteinase #status predicted <MAT>
 F:136-374/Domain: subtilisin homology <SBT>
 F:145,185,360/Active site: Asp, His, Ser #status predicted

Query Match 12.3%; Score 277; DB 1; Length 420;
 Best Local Similarity 31.2%; Pred. No. 1.9e-10;
 Matches 98; Conservative 44; Mismatches 120; Indels 52; Gaps 16;

QY 21 GGGQIVAVADTGLDGRNDSSMHEAFRGKITAL--YALGRT---NNANDTGHGTHVAGS 75
 DB 136 GGGINIAVLDGTGNTN-----HPDLNNVEOCKFTVGTITNNSCCTDRQGHGTHVAGS 189
 QY 76 VLNGSTNK---GMAPOANL-VFQSIMDSGGGLG-GLFSLNQLTFLPSQAYSAGARIHTN-S 129
 DB 190 ALADGGTGNVGVAPADADLWAYKVLGDDSGYADDIAAIRHAGDQATALNTKVINNS 249
 QY 130 WGAANVGATTSRNVDDYVRKNDMTILPAAAGNEGNGGTISAPGTAKNAITVGATENLR 189
 DB 250 LGSSESSITNAVN---YSYKNGVLIIIAAGNSGPGYQSGIGYPGALVNAVAVALEN-K 305
 QY 190 PFGSGYADNINHVAQFSSRGPT-KDG-----RIKPDVMAPGTFILSARSSLAPDSSFVAN 243
 DB 306 VENGTY-----RVADFSSRGYSWTGDYAIQKGDVEISAPGAIIYST-----N-- 348
 QY 244 HDSKIAYMGSTWATPIVAGNVAQLREHFVKRGITPKPSLLKALIAAGADIGLGPNG 303
 DB 349 FDGGYATISGTSNASPHAAGLAAKTIWAQYSPASNVDRGELQYRAY---ENDILSGYYAG 405
 QY 304 -----NQGGGRVTL 312
 DB 406 YGDDFASGFGFATV 419

RESULT 12
 JC4908
 alkaline serine proteinase (EC 3.4.-.-) I precursor - Alteromonas sp.
 A:Alternate names: subtilase
 C:Species: Alteromonas sp.
 C:Date: 01-Nov-1996 #sequence_revision 01-Nov-1996 #text_change 08-Oct-1999
 C:Accession: JC4908
 F:tsujiibo, H.; Miyamoto, K.; Tanaka, K.; Kaidzu, Y.; Imada, C.; Okami, Y.; Inamori, Y.
 Biosci. Biotechnol. Biochem. 60, 1284-1288, 1996
 A:Title: Cloning and sequence analysis of a protease-encoding gene from the marine bacterium *Alteromonas*
 A:Reference number: JC4908; MUID:97141200; PMID:8987544
 A:Accession: JC4908
 A:Status: Preliminary
 A:Molecule type: DNA
 A:Cross-references: DDBJ:D38600; NID:G1536787; PIDN:BAAL8912.1; PID:d1019647; PID:g21602
 A:Experimental source: strain O-7
 C:Comment: This enzyme belongs to class I subtilisin-like family. It is a chelator-sensitized subtilisin.
 C:Genetics:
 A:Gene: aprI
 C:Superfamily: subtilisin homology
 C:Keywords: hydrolase
 F:1-40/Domain: signal sequence #status predicted <SIG>
 F:41-150/Domain: amino-terminal propeptide #status predicted <ATP>
 F:151-496/Product: alkaline serine protease
 F:182-452/Domain: subtilisin homology <SBT>
 F:497-715/Domain: carboxyl-terminal propeptide #status predicted <CTP>
 F:739-294,335-372,478-481/Disulfide bonds: #status predicted

Query Match 11.8%; Score 265; DB 2; Length 715;
 Best Local Similarity 25.8%; Pred. No. 2.2e-09;
 Matches 124; Conservative 48; Mismatches 170; Indels 138; Gaps 22;

QY 21 GGGQIVAVADTGLDGRNDSSMHEAFRGKITALYALGRTNAND-----TNG---- 67
 DB 182 GGGVAVAVLDGTGRPHLDLDANILPGYDMISNTFVANDGGARDNDARDPQDVAITRECGT 241
 QY 68 -----HGTHVAG-----SVLNGSGTNKGMAPQANLVFQSIMDSGGGLGLP 108

RESULT 13
 A72647
 Probable surface layer-associated STABLE proteinase APE0607 - Aeropyrum pernix (strain A72647)
 C:Species: Aeropyrum pernix
 C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999
 C:Accession: A72647
 R:Kawarayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takawa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; DNA Res. 6, 83-101, 1999
 A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, *Aeropyrum*
 A:Reference number: A72450; MUID:99310339; PMID:10382966
 A:Accession: A72647
 A:Status: Preliminary
 A:Molecule type: DNA
 A:Residues: 1-1331 <RAW>
 A:Cross-references: DDBJ:AF000060; NID:G5104188; PIDN:BAA79577.1; PID:d1043363; PID:951
 A:Experimental source: strain K1
 C:Genetics:
 A:Gene: APE0607

Query Match 11.7%; Score 264; DB 2; Length 1331;
 Best Local Similarity 23.6%; Pred. No. 6e-09;
 Matches 111; Conservative 66; Mismatches 158; Indels 136; Gaps 18;

QY 8 VKADVAOSSGLYCGQGVAVADTGLDGRNDSSMHE-AFRGKITALYA----- 55
 DB 346 VYADLS-TAYLFP-----LKALSDTGMTSGEPDPSLLDLSFADETPASYGSEVLARDFTGD 400
 QY 56 -----LGRTNAN-----DTNGH 68
 DB 401 GVNDFSAGALAGTYDWMVGLLTGSEVNLGRLGFDYAGLVPLGLDPQGRWVSLYDTLAH 450
 QY 69 GTHVAGSVLNGSTN-----KGMAPQANLVFQSIMDSGGGLGLPSNLQTLFSQA 118
 DB 461 GTSVATVIAIRGNVEFNLGYIETSLRGAFAKIA-----AGGSF-----LINVFAQL 509
 QY 119 YSAG-----ARIHTNSWG-----AANVGAYT---TDSRNVDDY-VRKNDMT 155
 DB 510 FLSGFEPQDPLNVMVYTGHEQVDVNNWNSVIALRGFLTGADYATIEDYIVSAGTV 569
 QY 156 ILFAAGNEGNGGTISAPGTAKNAITVGATE--NLRFSGSYADNINHVAQFSSRGPTKD 213
 DB 570 IVHAMNGGFOYGTATTPGAGSLIISVCASTLFDYRFFYGLPSPGCDVLSWDRGFSQI 629
 QY 214 GRIKPDVMAPGTFILSARSSLAPDSSFVANHDSKIAYMGSTWATPIVAGNVAQLREHFV 273

Db 630 GVAKPDVNTGSAWAG----VPVLTGLNGSLAFDFGCTSEATPTMTSGSVALVISAYQ 685
Qy 274 KYRGITEPKSLKAAIAGAADIGLGYPNGNQGWGRVTLDKSLNV-----AY 320
Db 686 QAFGAKSPGLVKAILAKSTARDTGA--DAFTQSGGVVYRAVKAVLGGVPIALSTSYV 743
Qy 321 VNSSSLSTSQKATYSTATAGKPLKLSLWSDA--PASTASVTLVNDLN 369
Db 744 ENVYSLLS-----GYSYFFLAPNVEDTQYPGVLKPEGTAVETLVKLTL 789
RESULT 14
A35742
aqualysin (EC 3.4.21.-) I precursor - Thermus aquaticus
C:Species: Thermus aquaticus
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 16-Jun-2000
C:Accession: A35742; S00620; S00324
R:Terada, I.; Kwon, S.T.; Miyata, Y.; Matsuzawa, H.; Ohta, T.
J. Biol. Chem. 265, 6576-6591, 1990
A:Title: Unique precursor structure of an extracellular protease, aqualysin I, with NH-2
A:Reference number: A35742; MUID:90216674; PMID:2182621
A:Accession: A35742
A:Molecule type: DNA
A:Residues: 1-513 <TER>
A:Cross-references: GB:J04108; GB:D90108; NID:g217171; PIDN:BAAL4135.1; PID:9
A:Note: the authors translated the codon CTG for residue 470 as Val, and GGT for residue
R:Kwon, S.T.; Terada, I.; Matsuzawa, H.; Ohta, T.
Eur. J. Biochem. 173, 431-437, 1988
A:Title: Nucleotide sequence of the gene for aqualysin I (a thermophilic alkaline serine
A:Reference number: S00620; MUID:8825062; PMID:3286255
A:Accession: S00620
A:Molecule type: DNA
A:Residues: 75-442 <KWO>
A:Cross-references: EMBL:X07734; NID:g48069; PIDN:CAA30559.1; PID:g602091
A:Note: part of this sequence, including the amino and carboxyl ends of the mature prote
R:Matsuzawa, H.; Tokugawa, K.; Hamaki, M.; Mizoguchi, M.; Taguchi, H.; Terada, I.; Kwon
Eur. J. Biochem. 171, 441-447, 1988
A:Title: Purification and characterization of aqualysin I (a thermophilic alkaline serin
A:Reference number: S00324; MUID:88151937; PMID:3162211
A:Accession: S00324
A:Molecule type: protein
A:Residues: 128-170 <MATS>
C:Superfamily: subtilisin; subtilisin homology
C:Keywords: extracellular protein; hydrolase; serine proteinase
F:1-14/Domain: signal sequence #status predicted <SIG>
F:15-127/Domain: propeptide #status predicted <PRO>
F:128-408/Product: aqualysin I #status experimental <MAT>
F:157-364/Domain: subtilisin homology <SBT>
F:255-257,281-283/Region: SI specificity crevice #status predicted
F:409-513/Domain: carboxyl-terminal propeptide #status predicted <CPR>
F:166,197,349/Active site: Asp, His, Ser #status predicted
Query Match 11.7%; Score 262.5; DB 1; Length 513;
Best Local Similarity 26.2%; Pred. No. 2.1e-09;
Matches 118; Conservative 43; Mismatches 144; Indels 145; Gaps 23;
Qy 16 SYGLYGQGVAVADTGLDTRNDSSMHEAFRGIYALYALG 74
Db 152 TYTAGRGVNVYVDTGIRT-----THREFGGRARVGVYDALG--GNGDCCNGHGHVAG 203
Qy 75 SVLNGSTNGKMAPOANLVFQSIMD--SGGLGLPSNLQTLFSQAYSAGARHTN---- 128
Db 204 TI---GGVTVGVAKVNLVAVRLDCCNGSGSTSGVAGVDWV-----TRNRRPAPA 252
Qy 129 --SWGAAVNGAYTDSRNVDYVRKN---DMTILFAAGNEGPNGGTISAPGTAKNAITVG 183
Db 253 NMSLGGGVSTA-----LDNAVKNISIAAGVYVAAGNDNACNYS-PARVAEALTVG 304
Qy 184 AT--ENLRPSFGSYADNINHVAFSSRGPTKGRIPDVMAPGTFILSARSSLAPSSFW 241
Db 305 ATTSSDARASFSNYGSCV-----DLFAPGASIPSA-----W 335

Qy 242 ANHDSKYAMGSGTSMATPIVAGNVAQLREHFVKRGTTP-----KPSLLKAAIAGAADIGL 298
Db 336 YTSDTATQTLNGTSMATPHVAG--VAAL--YLEQNPSATPASFASAILNGATTGLSLGGIS 392
Qy 299 GYPNGNQGWRTLDKSLNVAVNVNESSLSLSQKATYSTATAGKPLKLSLWSDAPAST 358
Db 393 GSPN-----RLLYSLSSGS-----GSTAPCTS 415
Qy 359 ---TASVTLVNDLNLVITAPNGTQY-----VGNDFTSPTSYNDNWGRN--- 397
Db 416 CSYITGSLSGPDYNF---QPNGTVYYSYPAGTHRAWLRGPAGTDF-DLYLWRWDGSRWLT 471
Qy 398 -----NVENVFINAPQSGTYTIEVQAYN 420
Db 472 VGSSTGPTSEESLSYSTAGTYLWRIYAYS 501
RESULT 15
C84120
subtilisin-type proteinase (EC 3.4.21.-) BH3763 precursor [similarity] - Bacillus halodi
C:Species: Bacillus halodurans
C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
C:Accession: C84120
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hir
Nucleic Acids Res. 28, 4317-4331, 2000
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
A:Reference number: A83650; MUID:20512582; PMID:11058132
A:Accession: C84120
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-757 <STO>
A:Cross-references: GB:AP001519; GB:BA000004; NID:g10176109; PIDN:BA07482.1; GSPDB:GN00
A:Experimental source: strain C-125
C:Genetics:
A:Gene: BH3763
C:Superfamily: microbial serine proteinase vpr; subtilisin homology
C:Keywords: hydrolase; serine proteinase
F:1-20/Domain: signal sequence #status predicted <SIG>
Query Match 11.4%; Score 256.5; DB 2; Length 757;
Best Local Similarity 22.2%; Pred. No. 8.5e-09;
Matches 117; Conservative 59; Mismatches 157; Indels 193; Gaps 19;
Qy 8 VKADVAQSSYGLYGQGVAVADTGLDTRNDSSMHEAFRG-----KITALYALG 57
Db 119 VRGMLDESGVHLTGKGVKVAVIDTGIDYTHPD--LQSSYKGYDFVDYDDPMETIASQG 176
Qy 58 RTNNANDNGHGTGVAGSVLNGSTNGKMAPOANLVFQSIMDSSGGLGLPSNLQTLFSQ 117
Db 177 -----PTLHGTVHVGIIAANGQV-KGVAPAEIYAYRALGPGG--QGTEQVIAAEK 227
Qy 118 AYSAGARIHTNSWGAANGAYTTDSRNVDYVRKNDMTILFAAGNEGPNGGTISAPGTAK 177
Db 228 AVEDGVDVNLGLNTVNGPDWPTSLADAAVEGVAVT--SNGNSGPNMWTGSPGTSK 286
Qy 178 NAITVGAT-----EN--LRPSFGS----- 194
Db 287 KAISVGASAPPLNTFTYLTAFGEENEISLYPPSGSLPWAFKRLDPMIDVGYGTEKEWGV 346
Qy 195 -----YADNINH----- 201
Db 347 AEGKVLIKRGWVPTEKVMHAAKARGVLIYNNTPGPTGMEGVNIPVWSITREDG 406
Qy 202 -----VAQSSRGP-TKGRIKPDVWAPGTFFILSARSS 233
Db 407 BFLLEQLQKNKELTLTIYRKEEDFVALFESSRGVPTHTWVDKPDVWAPGVSI-----DS 462
Qy 234 LAPDSSFWANHDSKYAVMGGTSMATPIVAGNVAQLREHFVKRGTTPK--PSLLKAAI 291
Db 463 TIPNNG-----YLGNGTSMAPHVAGAAAIKQ-----AHPEWPEQVKAALMN 507
Qy 292 GAADI-----GLGYNGNQGWRTLDKSLNVAVNVNESSLSLSQKATYSTATAGKPLKI 347

Db 508 TAKKLVQEGVPHEIHEQGAGRIQVDKAV-----AATSLVYFGALSFGK---- 551
QY 348 SLVWSDAPASTTASVTLVNDLNLVITAPNGTOYVGNDFTSPYNDNW 393
Db 552 ---WSKDDLREKRPVTLTIENHDTV---KRTYHISPPFDVDPDGVEM 591

Search completed: March 10, 2004, 14:55:11
Job time : 22 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2004, 14:47:28 ; Search time 17 Seconds
(without alignments)
1329.320 Million cell updates/sec

Title: US-09-985-689A-1-COPY

Perfect score: 2247

Sequence: 1 NDVARGIVKADVAQSSYGLY.....EVOQVNPVGPQTFSIAIWN 434

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Match | Length | DB ID | Description |
|------------|-------|-------|--------|-------|-------------|
| 1 | 519.5 | 23.1 | 1743 | 1 | TAGC_DICDI |
| 2 | 492 | 21.9 | 1905 | 1 | TAGB_DICDI |
| 3 | 311.5 | 13.9 | 806 | 1 | SUBV_BACSU |
| 4 | 305.5 | 13.6 | 1398 | 1 | PLS_PYRFU |
| 5 | 278 | 12.4 | 580 | 1 | EXPR_XANCP |
| 6 | 277 | 12.3 | 420 | 1 | SUBT_BACS9 |
| 7 | 262.5 | 11.7 | 513 | 1 | AQLI_THEAQ |
| 8 | 253.5 | 11.3 | 894 | 1 | WPRB_BACSU |
| 9 | 246.5 | 11.0 | 402 | 1 | ALP_CEPAC |
| 10 | 244.5 | 10.9 | 534 | 1 | PROA_VIBAL |
| 11 | 242.5 | 10.8 | 401 | 1 | THES_BACSP |
| 12 | 240 | 10.7 | 269 | 1 | SUBS_BACLE |
| 13 | 240 | 10.7 | 380 | 1 | ELYA_BACAO |
| 14 | 240 | 10.7 | 380 | 1 | ELYA_BACCS |
| 15 | 239 | 10.6 | 269 | 1 | PRTM_BACSP |
| 16 | 235.5 | 10.5 | 382 | 1 | SUBT_BACAM |
| 17 | 235 | 10.5 | 378 | 1 | ELYA_BACSP |
| 18 | 232 | 10.3 | 269 | 1 | SUBB_BACLE |
| 19 | 231 | 10.3 | 321 | 1 | ISP_BACCS |
| 20 | 230.5 | 10.3 | 379 | 1 | SUBT_BACLI |
| 21 | 230 | 10.2 | 404 | 1 | SMPI_MAGFO |
| 22 | 227 | 10.1 | 1181 | 1 | SCA2_STRPY |
| 23 | 226.5 | 10.1 | 1052 | 1 | MS1P_CRIGR |
| 24 | 226.5 | 10.1 | 1052 | 1 | MS1P_HUMAN |
| 25 | 226.5 | 10.1 | 1052 | 1 | MS1P_MOUSE |
| 26 | 226.5 | 10.1 | 1052 | 1 | MS1P_RAT |
| 27 | 225.5 | 10.0 | 381 | 1 | SUBN_BACNA |
| 28 | 225.5 | 10.0 | 1167 | 1 | SCAL_STRPY |
| 29 | 223.5 | 9.9 | 381 | 1 | SUBT_BACSA |
| 30 | 223.5 | 9.9 | 381 | 1 | SUBT_BACST |
| 31 | 223.5 | 9.9 | 381 | 1 | SUBT_BACSU |
| 32 | 221.5 | 9.9 | 275 | 1 | SUBT_BACPU |
| 33 | 216.5 | 9.6 | 274 | 1 | SUBD_BACLI |

| | | | | | |
|----|-------|-----|------|---|------------|
| 34 | 216.5 | 9.6 | 530 | 1 | HLV_HALI7 |
| 35 | 216.5 | 9.6 | 645 | 1 | SUBB_BACSU |
| 36 | 215 | 9.6 | 422 | 1 | TKSU_PYRKO |
| 37 | 214.5 | 9.5 | 361 | 1 | ELYA_BACHD |
| 38 | 214 | 9.5 | 279 | 1 | THET_THEVU |
| 39 | 214 | 9.5 | 293 | 1 | PRTT_TRIAL |
| 40 | 214 | 9.5 | 326 | 1 | ISP_FAEPO |
| 41 | 214 | 9.5 | 409 | 1 | ALP_TRIHA |
| 42 | 211 | 9.4 | 319 | 1 | ISPI_BACSU |
| 43 | 210.5 | 9.4 | 1433 | 1 | SUBF_BACSU |
| 44 | 208.5 | 9.3 | 388 | 1 | CUDP_METAN |
| 45 | 208 | 9.3 | 387 | 1 | PRTR_TRIAL |

ALIGNMENTS

| | | | | | |
|------------|--|-----------|------|------|-----|
| RESULT 1 | | | | | |
| TAGC_DICDI | | | | | |
| ID | TAGC_DICDI | STANDARD; | PRT; | 1743 | AA. |
| AC | Q23868; | | | | |
| DT | 01-NOV-1997 (Rel. 35, Created) | | | | |
| DT | 01-NOV-1997 (Rel. 35, Last sequence update) | | | | |
| DT | 15-MAR-2004 (Rel. 43, Last annotation update) | | | | |
| DE | Prestalk-specific protein tagC precursor (EC 3.4.21.-). | | | | |
| GN | TAGC. | | | | |
| OS | Dictyostelium discoideum (slime mold). | | | | |
| OC | Eukaryota; Mycetozoa; Dictyostellida; Dictyostelium. | | | | |
| OX | NCBI_TaxID=44689; | | | | |
| RN | [1] | | | | |
| RP | SEQUENCE FROM N.A. | | | | |
| RC | STRAIN=AX4; | | | | |
| RX | MEDLINE=97140317; PubMed=8986798; | | | | |
| RA | Shaulsky G., Escalante R., Loomis W.F.; | | | | |
| RT | "Developmental signal transduction pathways uncovered by genetic suppressors"; | | | | |
| RL | Proc. Natl. Acad. Sci. U.S.A. 93:15260-15265 (1996). | | | | |
| CC | -!- FUNCTION: Intercellular communication via tagC may mediate integration of cellular differentiation with morphogenesis (By similarity). | | | | |
| CC | -!- SIMILARITY: In the N-terminal section; belongs to peptidase family S8. | | | | |
| CC | -!- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE ATP-BINDING TRANSPORT PROTEIN FAMILY (ABC TRANSPORTERS). MDR SUBFAMILY. | | | | |
| CC | -!- SIMILARITY: STRONG, TO TAGC. | | | | |
| CC | ----- | | | | |
| CC | This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch). | | | | |
| CC | ----- | | | | |
| DR | EMBL; U60086; AAB03331.1; -- | | | | |
| DR | PIR; T18279; T18279. | | | | |
| DR | DictyBase; DDB0001795; tagC. | | | | |
| DR | InterPro; IPR003593; AAA ATPase. | | | | |
| DR | InterPro; IPR001140; ABC_TM_transpt. | | | | |
| DR | InterPro; IPR003439; ABC_transporter. | | | | |
| DR | InterPro; IPR00209; Peptidase_S8. | | | | |
| DR | Pfam; PF00664; ABC_membrane; 1. | | | | |
| DR | Pfam; PF00005; ABC_tran; 1. | | | | |
| DR | Pfam; PF00082; Peptidase_S8; 1. | | | | |
| DR | PRINTS; PR00723; SUBTILISIN. | | | | |
| DR | SMART; SM00382; AAA_1 | | | | |
| DR | PROSITE; PS00929; ABC_TM1P; 1. | | | | |
| DR | PROSITE; PS00211; ABC_TRANSPORTER_1; 1. | | | | |
| DR | PROSITE; PS00893; ABC_TRANSPORTER_2; 1. | | | | |
| DR | PROSITE; PS00136; SUBTILASE_ASP; FALSE_NEG. | | | | |
| DR | PROSITE; PS00137; SUBTILASE_HIS; 1. | | | | |
| DR | PROSITE; PS00138; SUBTILASE_SER; FALSE_NEG. | | | | |
| KW | Hydrolase; Serine protease; ATP-binding; Transport; Transmembrane; | | | | |

FT SIGNAL 1 31
 FT CHAIN 32 1905
 FT DOMAIN 378 700
 FT DOMAIN 1518 1756
 FT TRANSMEM 1011 1031
 FT TRANSMEM 1076 1096
 FT TRANSMEM 1121 1141
 FT TRANSMEM 1210 1230
 FT TRANSMEM 1309 1329
 FT TRANSMEM 1332 1352
 FT ACT_SITE 387 387
 FT ACT_SITE 432 432
 FT ACT_SITE 695 695
 FT NP_BIND 1553 1560
 FT DOMAIN 63 67
 FT DOMAIN 95 104
 FT DOMAIN 107 134
 FT DOMAIN 311 321
 FT DOMAIN 833 837
 FT DOMAIN 838 844
 FT DOMAIN 871 876
 FT DOMAIN 1012 1015
 FT DOMAIN 1386 1389
 FT DOMAIN 1398 1404
 FT DOMAIN 1445 1450
 FT DOMAIN 1765 1779
 FT DOMAIN 1782 1785
 FT DOMAIN 1807 1812
 FT DOMAIN 1813 1860
 FT DOMAIN 1872 1878
 FT CARBOHYD 594 594
 FT CARBOHYD 621 621
 FT CARBOHYD 672 672
 FT CARBOHYD 747 747
 FT CARBOHYD 823 823
 FT CARBOHYD 1172 1172
 FT CARBOHYD 1522 1522
 FT CARBOHYD 1858 1858
 FT CARBOHYD 1905 1905
 FT SEQUENCE 21.9%; Score 492; DB 1; Length 1905;
 Best Local Similarity 27.9%; Pred. No. 2.7e-24;
 Matches 161; Conservative 73; Mismatches 162; Indels 182; Gaps 22;

QY 19 LYGGQIVAVADTGLDTCR-----NDS-----SMHEAFRGKITALYALGRNNDANTNGH 68
 DB 376 LRKGQILSIADTGLDGHCHCFSDSKYPIPNQVNHNRKVVT---YITHDNEYVNGH 432
 QY 69 GTHVAGSVLNG-----STNKGWAPQANLVFQSIDSGGLGLPSNLQTLFQAYS 121
 DB 433 GTHVCGSAAGTPEDSWAISSFSGLATDAFYD-LSSGSSEPTPEDYSQMKPLYDA 491
 QY 122 GARTHSWGA-----AVNGATTSRNVDDYVRK-NDMTILFAAGNEGPGGTISAPGTA 176
 DB 492 GARVHGSWSGVSQGYGYGSDDAGGIDAFLEYEPFSILRAAGN-NELFASLLAQATA 550
 QY 177 KNATITVGATENLRPSFGS-----YADNI-----199
 DB 551 KNAITVGAETAHVYVSDALEYDFSDNANFORPCLFDKXCNYYTAKCSESVNVKGL 610
 QY 200 -----NHVAFSSRGPTKGRIKPDVNAAPTFTILSARSS- 233
 DB 611 QLCCPASIKQNASDFTTQPOFYNNENNGSFSKGPHTDGLKPKDIPVAPGEYITSARSNG 670
 QY 234 -----LAPDSFWANHDSKYAVMGSTMAPIVAGNVAQLREHF-----VKNRGI 278
 DB 671 ENSTDQCGDGL--PNANGLMSISGTNVAIPLATAATILRQVLDVGYFTGSEVENKL 728
 QY 279 TPKEPSLKAALIAGADIGLY-----PNGNQGWGRTLDKSLNVA 319
 DB 729 LPTGSLKALMINNAQLNGTYFWFASGTSNPSNAIFQINGANLIQGWGALRWN---NWL 785
 QY 320 YVNESS-----SLSTSQKAT-----YST-ATAGK 343

DB 786 YKSSNPFPSPRWIGIGLGNKQKATEWKEDSLSLGKNSCYFTYKPSSSSGSGGGGT 845
 QY 344 P-LKISLVMSDAPATTASVTLVNDLNL-----VITAPN--GTQYVGNDFSTPYND 391
 DB 846 PRIVATLVMTDPPSYSGAKFNLVNNDLILLNSDDDSITIGSGGSLQPAKGVAQP--- 902
 QY 392 NWDGNNVENVFINAPOSGTGTIEVOAVNPVPGPOTFS 429
 DB 903 --DTLNNVEGIIINFTKAMNYKFTIAGTNVPIGPQKFS 938

RESULT 3
 SUBV_BACSU STANDARD; PRT; 806 AA.
 AC P29141.
 DT 01-DEC-1992 (Rel. 24, Created)
 DT 01-DEC-1992 (Rel. 24, Last sequence update)
 DE 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Minor extracellular protease vpr precursor (EC 3.4.21.-).
 GN VPR OR IPA-45R OR BSU38090.
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 OX NCBI_TaxID=1423;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 161-195.
 RX MEDLINE=92041574; PubMed=1938892;
 RA Sloma A., Rufo G.A. Jr., Theriault K.A., Dwyer M., Wilson S.W.,
 Pero J.;
 RT "Cloning and characterization of the gene for an additional
 extracellular serine protease of *Bacillus subtilis*."
 RL J. Bacteriol. 173:6889-6895(1991).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RX MEDLINE=95020337; PubMed=7934828;
 RA Glaser P., Kunst F., Arnaud M., Coudart M.P., Gonzales W.,
 Hullo M.F., Ionescu M., Lubochinsky B., Marcelino L., Moser I.,
 Presecan E., Santanu M., Schneider E., Schweizer J., Verzes A.,
 Rapoport G., Danchin A.;
 RT "Bacillus subtilis genome project: cloning and sequencing of the 97
 kb region from 325 degrees to 333 degrees."
 RL Mol. Microbiol. 10:371-384(1993).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RX MEDLINE=98044033; PubMed=9384377;
 RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
 Azevedo V., Bortaro M.G., Bessieres P., Bolotin A., Borchert S.,
 Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
 Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
 Chou S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
 Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
 Entian K.D., Brington J., Fabret C., Ferrari E., Foulger D.,
 Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
 Ghim S.Y., Glaser P., Goffeau A., Gollightly E.J., Grand G.,
 Guisepi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
 Hilbert H., Holsappel S., Hosono S., Hullo M.P., Itaya M., Jones L.,
 Joris B., Karamata D., Kashiwara Y., Klaerr-Blanchard M., Klein C.,
 Kobayashi Y., Koster P., Koningsstein G., Krogh S., Kumano M.,
 Kurita K., Lapidus A., Lardin S., Lauber J., Lazarevic V.,
 Lee S.M., Levine A., Liu H., Masuda S., Maue C., Medigue C.,
 Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
 Noone D., O'Reilly M., Ogawa K., Ogiwara K., Oudega B., Park S.H.,
 Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
 Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
 Rieger M., Rivolta C., Rocha S., Roche S., Rose M., Sadate Y.,
 Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
 Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
 Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
 Takeuchi M., Tamakoshi A., Tanaka T., Terpestra P., Tognoni K.,
 Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
 Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T.,


```
DR EMBL; AB012184; AAM40166.1; -.
DR PIR; S11890; S11890.
DR HSP; P00782; 2SPT.
DR MEROPS; S08.004; -.
DR InterPro; IPR000209; Peptidase_S8.
DR InterPro; IPR007280; PPC.
DR InterPro; IPR009020; Protease_inhib.
DR Pfam; PF00082; Peptidase_S8; 1.
DR PRINTS; PR00723; SUBTILISIN.
DR PROSITE; PS00136; SUBTILASE_ASP; 1.
DR PROSITE; PS00137; SUBTILASE_HIS; 1.
DR PROSITE; PS00138; SUBTILASE_SER; 1.
KW Hydrolase; Serine protease; Zymogen; Signal; Complete proteome.
FT SIGNAL 1 32
FT PROPEP 33 7136
FT CHAIN 7137 580
FT ACT_SITE 177 177
FT ACT_SITE 237 237
FT ACT_SITE 409 409
FT DISULFID 225 273
FT DISULFID 315 352
FT DISULFID 450 454
SQ SEQUENCE 580 AA; 57228 MW; 8C9A2CAE4E7F47CB CRC64;

Query Match 12.4%; Score 278; DB 1; Length 580;
Best Local Similarity 26.0%; Pred. No. 5.5e-11;
Matches 126; Conservative 56; Mismatches 147; Indels 156; Gaps 24;

QY 21 GQGIIVAVADTGL-----DTRNDSSMHEAFRGKITALYALGRNNAND-----64
DB 168 GSGTVVAVIDTGTSHADLNANLAGYDFISDATTARDNGRDSNADEGDIYAAANECA 227
QY 65 -----TNGHGTTHVAGSLVNGSTNGKMAPOA-----NLVFQSIMD 99
DB 228 GIPAASSWHGTHVAGTVAATNTTGVAGTAGAKVVPVVLKCGGSLDIADIVA 287
QY 100 SGGGLGGLPSNLO--TLFQAYAGARIHTNSWGAANGAYTTDSRNVDDYVRKNDMTIL 157
DB 288 SGGTVSGIPANAPAEVINNSLGGGSCSTTMQN-AINGAVSRGT-----TV 334
QY 158 FAAGNPGNGGTTSAPOTAKNAVITVGATEN--LRPSFGSYADNINHVAFSSRGPTKGR 215
DB 335 VAAGNDASVSG-SLPANCANVIAVAATTSAGKAKASYNFTGTI-----377
QY 216 IKPDVMPAGPFLISARSS--LAPDSFWANHDSKIYMGGTSMATPIVAGNAQLRHFV 273
DB 378 ---DVSAGPSSILSTLNSGTTTPGS-----ASYASVNGTSMASPHVAGVALVQS--V 425
QY 274 KNRGITPK--PSLLK--AALIAGAADIGLYPNNGOGWRVTLDKSLNVAYNSS----325
DB 426 APTALTPAAVEILLKNTARALPAC-----SGCGAGIVNADAATVA--INGSGGGG 477
QY 326 -----SLTSQKATYGTATAGKPLKISLWSDAPASTTASVTL---VND 367
DB 478 GGGNTLTNGTPVTGLGAATGAELNYITITPAG-----SGTLTVTTSGGSD 523
QY 368 LNLVI---TAPNGTQVGNDFTPSYNDNDGRNVENVFAPNAPQSGTYTIEVCAYNPVG 424
DB 524 ADLYVAGSAPTSAYT---CRPYRS-----GNAETCTITAP-SGTYTVRLKAYS-----569
QY 425 PQTFSS 429
DB 570 --TFS 572

RESULT 6
SUBT_BAC99
ID SUBT_BAC99 STANDARD; PRT; 420 AA.
AC P28842;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
```

```
DE Subtilisin precursor (EC 3.4.21.62).
GN SUBI.
OS Bacillus sp. (strain TA39).
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=29336;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92256481; PubMed=1591352;
RA Narinx E., Davail S., Feller G., Gerday C.;
RT "Nucleotide and derived amino acid sequence of the subtilisin from
RL the antarctic psychrotroph Bacillus TA39."
RL Biochim. Biophys. Acta 1131:111-113(1992).
CC -!- FUNCTION: Subtilisin is an extracellular alkaline serine protease,
CC it catalyzes the hydrolysis of proteins and peptide amides.
CC -!- CATALYTIC ACTIVITY: Hydrolysis of proteins with broad specificity
CC for peptide bonds, and a preference for a large uncharged residue
CC in P1. Hydrolyzes peptide amides.
CC -!- COFACTOR: Binds 1 calcium ion per subunit (Potential).
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- MISCELLANEOUS: Still active at temperatures close to 0 degrees
CC Celsius, it has a marked heat lability.
CC -!- MISCELLANEOUS: Secretion of subtilisin is associated with onset of
CC sporulation, and many mutations which block sporulation at early
CC stages affect expression levels of subtilisin. However, subtilisin
CC is not necessary for normal sporulation.
CC -!- SIMILARITY: Belongs to peptidase family S8.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X62369; CAA44227.1; -.
DR PIR; S23407; S23407.
DR HSP; Q99405; IMPT.
DR InterPro; IPR000209; Peptidase_S8.
DR Pfam; PF00082; Peptidase_S8; 1.
DR PRINTS; PR00723; SUBTILISIN.
DR PROSITE; PS00136; SUBTILASE_ASP; 1.
DR PROSITE; PS00137; SUBTILASE_HIS; 1.
DR PROSITE; PS00138; SUBTILASE_SER; 1.
KW Hydrolase; Sporulation; Serine protease; Zymogen; Metal-binding;
KW Calcium-binding; Signal.
FT SIGNAL 1 26
FT PROPEP 27 111
FT CHAIN 112 420
FT ACT_SITE 145 145
FT ACT_SITE 182 182
FT ACT_SITE 360 360
FT METAL 115 115
FT METAL 154 154
SQ SEQUENCE 420 AA; 44086 MW; AB4F121BD32B26EC CRC64;
```

```
Query Match 12.3%; Score 277; DB 1; Length 420;
Best Local Similarity 31.2%; Pred. No. 4.2e-11;
Matches 98; Conservative 44; Mismatches 120; Indels 52; Gaps 16;

QY 21 GQGIIVAVADTGLDTRNDSSMHEAFRGKITAL--YALGRT---NNANDTNGHGTTHVAGS 75
DB 136 GGGINIAVLDTGVNTN-----HPDLNNVEQCKDFVTGTTTNNSTCTDRQGHGTHVAGS 189
QY 76 VLNGSGTNNK---GMAPQANL-VFQSIMDSGGGLG-GLPSNLQTLFSAQYAGARIHTN-S 129
DB 190 ALADGGTGNGVYGVAPADLWAYKVLGGDGGSGVADDIAAAIRHAGDQATALNTKVVNMS 249
QY 130 WGAAVNGAYTTDSRNVDYVRKNDMTILFAAGNEGPNNGGTISAPGTAKNAITVGCATENLR 189
DB 250 LGSSGESLITNAYN---YSYNGKGVLLIAAGNSGFPYQSGISYPGALVNAVAALEN-K 305
```

```
QY 190 PFGSYADNINHVAQESSRGPT-KDG-----RIKPDVWAPGTILSARSLAPDSFWAN 243
Db 306 VENGTY-----RVADSSRGYSWTDGYAIQKGVDEISAPGAAYST-----W-- 348
QY 244 HDSKIYMGTSMTATPIVAGNVAQLRHFVKNRGITPKPSLLKAALIAGAADIGLGYPNG 303
Db 349 FDGGYATISGTSMASPHAAGLAAKIWAQYPSAGNVDRGELQVRAY---ENDILSGYYAG 405
QY 304 -----NQGWGRVTL 312
Db 406 YGDDFASGFGFATV 419

RESULT 7
AOL1_THEAQ
ID _AOL1_THEAQ STANDARD; PRT; 513 AA.
AC P08594;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Aqualysin I precursor (EC 3.4.21.-).
GN PSTI.
OS Thermus aquaticus.
OC Bacteria; Deinococcus-Thermus; Deinococci; Thermales; Thermaceae;
OC Thermus.
OX NCBI_TaxID=271;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 15-23.
RX STRAIN=YTL1;
MEDLINE=90216674; PubMed=2182621;
RA Terada I., Kwon S.-T., Miyata Y., Matsuzawa H., Ohta T.;
RT "Unique precursor structure of an extracellular protease, aqualysin
RT I, with NH2- and COOH-terminal pro-sequences and its processing in
RT Escherichia coli.";
RL J. Biol. Chem. 265:6576-6581(1990).
RN [2]
RP SEQUENCE OF 75-442 FROM N.A., AND PARTIAL SEQUENCE.
RX STRAIN=YTL1;
MEDLINE=86235062; PubMed=3286255;
RA Kwon S.-T., Terada I., Matsuzawa H., Ohta T.;
RT "Nucleotide sequence of the gene for aqualysin I (a thermophilic
RT alkaline serine protease) of Thermus aquaticus YT-1 and
RT characteristics of the deduced primary structure of the enzyme.";
RL Eur. J. Biochem. 173:491-497(1988).
RN [3]
RP SEQUENCE OF 128-170.
RX MEDLINE=88151937; PubMed=3162211;
RA Matsuzawa H., Tokugawa K., Hamaoki M., Mizoguchi M., Taguchi H.,
RA Terada I., Kwon S.-T., Ohta T.;
RT "Purification and characterization of aqualysin I (a thermophilic
RT alkaline serine protease) produced by Thermus aquaticus YT-1.";
RL Eur. J. Biochem. 171:441-447(1988).
CC -!- FUNCTION: Aqualysin I is a thermophilic alkaline serine protease.
CC the optimal temperature for its caseinolytic activity is 80
CC degrees Celsius.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- DEVELOPMENTAL STAGE: Secreted from the early stationary phase
CC until the time the cells cease to grow.
CC -!- PTM: THE N- AND C-TERMINAL PRO-SEQUENCES ARE REMOVED THROUGH THE
CC PROTEOLYTIC ACTIVITY OF AQUALYSIN I ITSELF, IN THAT ORDER. THE
CC C-TERMINAL PRO-SEQUENCE IS REQUIRED FOR TRANSLOCATION OF THE
CC PROTEASES ACROSS THE OUTER MEMBRANE.
CC -!- PTM: Two disulfide bonds are present.
CC -!- SIMILARITY: Belongs to peptidase family S8.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
```

```
DR EMBL; D90108; BAA14135.1; -.
DR EMBL; X07734; CAA30559.1; -.
DR FIR; A35742; A35742.
DR HSSP; P06873; 2PKK.
DR InterPro; IPR000209; Peptidase S8.
DR InterPro; IPR009020; Protease inhib.
DR Pfam; PF00082; Peptidase S8; 1.
DR PRINTS; PR00723; SUBTILISIN.
DR PROSITE; PS00136; SUBTILASE_ASP; 1.
DR PROSITE; PS00137; SUBTILASE_HIS; 1.
DR PROSITE; PS00138; SUBTILASE_SER; 1.
KW Hydrolase; Serine protease; Zymogen; Signal.
FT SIGNAL 1 14
FT PROPEP 15 127
FT CHAIN 128 408 AQUALYSIN I.
FT PROPEP 409 513
FT ACT_SITE 166 166 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 197 197 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 349 349 CHARGE RELAY SYSTEM (BY SIMILARITY).
SQ SEQUENCE 513 AA; 53913 MW; DDFDFE6D4A50B785 CRC64;

Query Match 11.7%; Score 262.5; DB 1; Length 513;
Best Local Similarity 26.2%; Pred. No. 4.9e-10; Indels 145; Gaps 23;
Matches 118; Conservative 43; Mismatches 144;

QY 16 SYGLYGGQIVAVADTGLDTGRNDSMHEAFRGKITALY-ALGRTNANDTNGHGTTHVAG 74
Db 152 TVTATGRGVNVYVDTGIRT-----THREFGGRARVGYDALG--GNGQDCNCGHGTTHVAG 203
QY 75 SVLGNSTNKGMAPQANLVFQSIIND--SGGGLGSLPSNLQTLFQAYSAGARIHTN---- 128
Db 204 TI---GGVTYGVAKAVNLVAVRVLDCNGSGTSGVIAGVDWV-----TRNRRRPAVA 252
QY 129 --SWGAAVNGAYTTDSRNVDVYRKN---DMTILFAAGNEGPNGGTISAPGTAKNAITVG 183
Db 253 NMSLGGGVSTA-----LDNAVKNSTAAGVYVAAAGNDNANACNYS-PARVAEALTVG 304
QY 184 AT--ENLRPFGSYADNINHVAQESSRGPTKGRIKPDVWAPGTILSARSLAPDSFW 241
Db 305 ATTSDDARASFSNYGSCV-----DLFAPGASIFSA-----W 335
QY 242 ANHDSKYAVMGGTSMATPIVAGNVAQLRHFVKNRGITP--YLEQNPSATPASVAILNGATTGRLSGIGS 298
Db 336 YTSATATQTLNGTSMATPHVAG-VAAL--YLEQNPSATPASVAILNGATTGRLSGIGS 392
QY 299 GYPNGNQGWGRVTLDKSLINVAYNVNESSLSQKATSYFTATAGKPLKISLVWSDAPST 358
Db 393 GSPN-----RLYSLISLSSGS-----GSTAPCTS 415
QY 359 ---TASVTLVNDLNLVITAPNGTOY-----VGNDFTSFYNDNWDGRN--- 397
Db 416 CSYVTGSLSGPDDYNF---QPNGTYYSYPAGTHPAWLRGAGTDF-DLYLWRWDGSRWLT 471
QY 398 -----NVENVFINAPQSGTYTIEVOAYN 420
Db 472 VGSSTGPTSEBSLSYSGTAGYLLRWIRIYAYS 501

RESULT 8
WPA_BACSU
ID WPA_BACSU STANDARD; PRT; 894 AA.
AC P54423; O06726;
DT 01-OCT-1996 (Rel. 34, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Cell wall-associated protease precursor (EC 3.4.21.-) [Contains: Cell
DE wall-associated peptidases CWBP23 and CWBP52].
GN WPA OR BSU10770.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
```

RP SEQUENCE FROM N.A., AND SEQUENCE OF 32-54 AND 414-428.
RC STRAIN=168;
RA MEDLINE=57158234; PubMed=9004506;
RX Margot P., Karamata D.;
RT "The wprA gene of *Bacillus subtilis* 168, expressed during exponential growth, encodes a cell-wall-associated protease.";
RL Microbiology 142:3437-3444(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA MEDLINE=98015415; PubMed=933931;
RX Medina N., Vannier F., Roche B., Autret S., Levine A., Seror S.J.;
RT "Sequencing of regions downstream of addA (98 degrees) and citG (289 degrees) in *Bacillus subtilis*.";
RL Microbiology 143:3305-3308(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Mosser I., Albertini A.M., Alloni G., Azevedo V., Betero M.G., Bessieres P., Bolotin A., Borchert S., Borriass R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S., Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M., Choi S.K., Codani J.J., Conerton I.F., Cummings N.J., Daniel R.A., Denizot F., Devine K.M., Duserchot A., Ehrlich S.D., Emerson P.T., Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D., Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N., Gim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G., Giuseppe G., Guy B.J., Haga K., Haeche J., Harwood C.R., Henaut A., Hilbert H., Holsappel S., Hosono S., Hulio M.F., Itaya M., Jones L., Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C., Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kunano M., Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V., Lee S.M., Levine A., Liu H., Masuda S., Maue C., Medigue C., Medina N., Mollato R.P., Mizuno M., Moestl D., Nakai S., Noback M., Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H., Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M., Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S., Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadia Y., Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F., Sekiguchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Soldo B., Sorokin A., Taccioni E., Takagi T., Takahashi H., Takemaru K., Takeuchi M., Tanakoshi A., Tanaka T., Terpstra P., Tognoni A., Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A., Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T., Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K., Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.";
RL Nature 390:249-256(1997).
CC -!- FUNCTION: NOT YET KNOWN; COULD BE INVOLVED IN PROTEOLYCAN DEGRADATION, BY CLEAVAGE OF ITS PEPTIDE BRIDGES.
CC -!- SUBCELLULAR LOCATION: Cell-wall bound.
CC -!- PTM: PROCESSED INTO CWBP23 AND CWBP52.
CC -!- SIMILARITY: Belongs to peptidase family 58.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U58981; AAC25926.1; -;
CC EMBL; Y09476; CAA70641.1; -;
CC EMBL; Z98109; CAB12917.1; -;
CC PIR; F69730; F69730.
CC HSP; Q45670; IDB1.
CC MEROPS; S08.004; -;
CC Subtilisin; BG11846; wprA.
CC InterPro; IPR000209; Peptidase_S8.
CC Pfam; PF00082; Peptidase_S8; 1.

DR PRINTS: PR00723; SUBTILISIN.
DR PROSITE; PS00136; SUBTILASE ASP; FALSE_NEG.
DR PROSITE; PS00137; SUBTILASE_HIS; 1.
DR PROSITE; PS00138; SUBTILASE_SER; 1.
KW Hydrolase; Serine protease; Cell wall; Zymogen; Signal;
KW Complete proteome.
FT SIGNAL 1 31 CELL WALL-ASSOCIATED PROTEASE.
FT CHAIN 32 894 CWBP23.
FT CHAIN 32 ? POTENTIAL.
FT PROPEP 2 413 CWBP52.
FT CHAIN 414 894 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 466 466 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 497 497 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 650 650 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT CONFLICT 9 9 V -> A (IN REF. 1).
FT CONFLICT 14 14 L -> I (IN REF. 1).
SQ SEQUENCE 894 AA; 96487 MW; 0F67C353E55F8DBC CRC64;
Query Match 11.3%; Score 253.5; DB 1; Length 894;
Best Local Similarity 24.7%; Pred. No. 3.9e-09;
Matches 108; Conservative 60; Mismatches 141; Indels 129; Gaps 19;
QY 25 IVAVADTGLDTRNDSSMHEAFRGKITALYA---LGRTNANDTNGHGTHTVAGSVLG--- 78
DB 457 LIAVVDITGVSTLAD-----LKGVRTDLGHNFVGRNNAMDDQGHGTHVAGIIAAQSD 510
QY 79 NGSTNKGMAQANLVFOSIMDSGGGLGGLPSNLQTLFSAQYASAGARIHTNSWGAAYNGAY 138
DB 511 NGYSMTGLNAKAKIIPVKVILDSAG--SGDTEQALGKIYAADKGAIVINLSLG-----GGY 564
QY 139 TTDSRNVDVVRKNDMTILFAAGNEGPNGGGTIGAPGAKNAITVGATENLRPSFGSYADN 198
DB 565 SRVLEFALKVAADKNVLIAAAGNDNGEN--ALSYPASSKYVMSVGT-----NR 611
QY 199 INHVAQSSRGRTKDKRIKPDVWAPGTFILSARSLAPDSFWANHDSKYAYMGTSMAT 258
DB 612 MDMTADFSTNYGKGL-----DISAPGSDI-----PSLVPNGN-----VTYMSGTSMAT 653
QY 259 PIVAGNVAQLREHFVKNGI--TPKPSLLKAALIAAGADIGLGPYNGNQ----- 305
DB 654 PYAAAAAGLL---FAQPKLKRTEVEDMLKT-----ADDISFESVDGGELELYDDYDPI 706
QY 306 -----GWRVTLDKSLNVA-----YVNESSLSSTSQKATYS----- 336
DB 707 EIPKTPGVDMHSGYGRNLNWKVSAADQLKVKNKLESTQTAVRGSAKEGTLIEVMNGKKK 766
QY 337 -FTATAGK--PLKISLVMSDAPASTTASVTLVNDLNLVITAPNGTQVGVNDFTSPYNDNW 393
DB 767 LGSAGAGKDNKAFKNI-----ATQKQDQVLYLKATG----- 798
QY 394 DGRNNVENVFINAPQSGT 411
DB 799 DAKTSYKVVVVKPKPSGT 816
RESULT 9
ALP_CEPAC STANDARD; PRT; 402 AA.
ID ALP_CEPAC
AC P29118;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Alkaline proteinase precursor (EC 3.4.21.-) (ALP).
GN ALP.
OS Cephalosporium acremonium (Acremonium chrysogenum).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Hypocreales; Hypocreaceae; Hypocreales; Hypocreales; Hypocreales;
OC Acremonium.
OC Acremonium.
OX NCBI_TaxID=5044;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9129283; PubMed=1368696;
RA Isogai T., Fukagawa M., Kojo H., Kohsaka M., Aoki H., Imanaka H.;

```

RT "Cloning and nucleotide sequences of the complementary and genomic
RT DNAs for the alkaline protease from Acremonium chrysogenum.";
RL Agric. Biol. Chem. 55:471-477(1991).
CC -1- SIMILARITY: Belongs to peptidase family S8.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; D00923; BAA00765.1; -
CC PIR; J03332; J03332.
CC HSSP; P06873; 2PRK.
CC MEROPS; S08.050; -.
CC InterPro; IPR000209; Peptidase_S8.
CC InterPro; IPR009020; Protease_inhib.
CC Pfam; PF00082; Peptidase_S8; 1.
CC PRINTS; PR00723; SUBTILISIN.
CC PROSITE; PS00136; SUBTILASE_HIS; 1.
CC PROSITE; PS00137; SUBTILASE_HIS; 1.
CC PROSITE; PS00138; SUBTILASE_SER; 1.
CC PROSITE; PS00139; SUBTILASE_SER; 1.
CC Hydrolase; Serine protease; Zymogen; signal.
CC SIGNAL 1 20 POTENTIAL.
CC FT CHAIN 121 402 ALKALINE PROTEINASE.
CC FT ACT_SITE 160 160 CHARGE RELAY SYSTEM (BY SIMILARITY).
CC FT ACT_SITE 191 191 CHARGE RELAY SYSTEM (BY SIMILARITY).
CC FT ACT_SITE 347 347 CHARGE RELAY SYSTEM (BY SIMILARITY).
CC SQ SEQUENCE 402 AA; 42099 MW; 8D030CCD42D918E1 CRC64;

Query Match 11.0%; Score 246.5; DB 1; Length 402;
Best Local Similarity 30.3%; Pred. No. 3.9e-09;
Matches 91; Conservative 32; Mismatches 102; Indels 75; Gaps 14;

QY 21 GGGQIVAVADTGLDTRNDSSMEAREGK-ITALYALGRNNDNTNGHGHVAGSVLGN 79
DB 151 GSGTYAVVDTGI-----LESHNEFSGRAITGVNAGGSN--ADINGHGHVAGTI--- 199

QY 80 GSTNKGMAPQANLQFQIMDSGGG-----LGLPSNLQTLFQAYAGARI-----HT 127
DB 200 GGRTYGVAKNTLTIKAVKFRGSSSTSIILDGFNVAVNDIINRGONKAAISMSLGGYVS 259

QY 128 NSWGAAYNGAYTDSRNVDDYVRKNDMTILFAAGNPGNGGTTISAPGTAKNAITVG--AT 185
DB 260 SAFNNAVNTAY---SRGVLSV-----AAGNDNQNAAYS-PASRAVAITVGSIAS 306

QY 186 ENLRPSFGSYADNINHVAFSSRGPTKDGRIKPDVMAPGTFFILSARSSLAPOSSFWANHD 245
DB 307 NWARSSPSNYGSQL-----DIFAPGTSILSA-----WIGGN 337

QY 246 SKYAYMGTSNATPIVAGNVAQLREHFVKNGRITPKPSLLKAALIAAGADI---GLGYPN 302
DB 338 SATNTISGTSNATPHVTGVVLYIQ-----ALEGLITSGAARINALATTGRVSNPSSGSPN 393

RESULT 10
PROA_VIBAL STANDARD; PRT; 534 AA.
AC P16588;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Alkaline serine exoprotease A precursor (EC 3.4.21.-).
GN PROA.
OS Vibrio alginolyticus.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrio.
CX NCBI_TaxID=663;
RN [1]
RP SEQUENCE FROM N.A.

```

```

RX MEDLINE=89326126; PubMed=2546861;
RA Deane S.M., Robb F.T., Robb S.M., Woods D.R.;
RT "Nucleotide sequence of the Vibrio alginolyticus calcium-dependent,
RT detergent-resistant alkaline serine exoprotease A.";
RL Gene 36:281-288(1989)
CC -1- SIMILARITY: Belongs to peptidase family S8.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M25499; AAA27550.1; -
CC PIR; J01173; J01173.
CC HSSP; Q99405; 1MPT.
CC MEROPS; S08.050; -.
CC InterPro; IPR000209; Peptidase_S8.
CC InterPro; IPR007280; PPC.
CC InterPro; IPR009020; Protease_inhib.
CC Pfam; PF00082; Peptidase_S8; 1.
CC Pfam; PF04151; PPC; 1.
CC PRINTS; PR00723; SUBTILISIN.
CC PROSITE; PS00136; SUBTILASE_HIS; 1.
CC PROSITE; PS00137; SUBTILASE_HIS; 1.
CC PROSITE; PS00138; SUBTILASE_SER; 1.
CC Hydrolase; Serine protease; Zymogen; signal.
CC SIGNAL 1 21 POTENTIAL.
CC FT CHAIN 142 534 ALKALINE SERINE EXOPROTEASE A.
CC FT ACT_SITE 180 180 CHARGE RELAY SYSTEM (BY SIMILARITY).
CC FT ACT_SITE 213 213 CHARGE RELAY SYSTEM (BY SIMILARITY).
CC FT ACT_SITE 363 363 CHARGE RELAY SYSTEM (BY SIMILARITY).
CC SQ SEQUENCE 534 AA; 55930 MW; 84E96D9C649D4226 CRC64;

Query Match 10.9%; Score 244.5; DB 1; Length 534;
Best Local Similarity 24.0%; Pred. No. 7.7e-09;
Matches 115; Conservative 69; Mismatches 138; Indels 157; Gaps 24;

QY 7 IVKADVAQSS--YGL-----YGGQIVAVADTGLDTRNDSSMEAREF 46
DB 137 IVSADANQTALWGLDIDQRNLPLDNNYSANFGTGTAYKIDTGV-----NNAIVEP 190

QY 47 RGTITALY-ALGRNNDNTNGHGHVAGSVLGNSTNKGMAPQANLQFQIM--DSGGG 103
DB 191 GGRSVSGYDFVNDADASDCNGHGHVAGTI--GGSLYGVAKNVLGVRLVSCSGSGS 247

QY 104 LGLPSNLQTLFQAYAGARIHTNSWGAAYNGAYTDSRNVDDYR---KNDMTILFAA 160
DB 248 TSGVIAGVDWAANA--SGPSVANMSLGGQGSVA-----LDSAVQSAVQSGVSFMLAA 298

QY 161 GNEGNGGTTISAPGTAKNAITVGAT--ENLRPSFGSYADNINHVAFSSRGPTKDGRIK 218
DB 299 GNSNADACNYS-PARVATGVTVGTTSTTDAESSFNNGSCV----- 338

QY 219 DVMAPGTFFILSARSSLAPOSSFWANHDSKYAYMGTSNATPIVAGNVAQLREHFVKRGI 278
DB 339 DVFAFGSQIKGA-----W-YDGGYKTIISGTSNATPHVAG-VAAL--YLOESSV 383

QY 279 TPKPSLLKAALIAAGADI-GLGYPNGNQNGRVTLDKSLNVAYNVNESSSLSSTSKATYSFT 338
DB 384 S--PSQVEALIVSRAST-----GKVT-----DTRGSVNKLLVSLT 416

QY 339 -----ATAGKPL-----KISLVMSDAPASTAGVTLVN----- 366
DB 417 DADCGQDCGPDPTDPPEGLKTSVPVSGSLSSGQAVYYVDVEAGORLTVQMYGSGSD 476

QY 367 -DLNLVITAPNGTQVGVNDFTSPYNDNWDGR-----NNVENVFINAPQSGTITIEQAVN 420
DB 477 ADLYLRFGA-----KPTLNAMDRCRPFKYGNNETCTVTSATQSGRTHVMIQGIS 523

```

```

RESULT 11
ID THES_BACSP STANDARD; PRT; 401 AA.
AC Q45670;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Thermophilic serine proteinase precursor (EC 3.4.21.-) (Ak.1
DE protease).
OS Bacillus sp. (strain AK1).
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1409;
RN [1]
SEQUENCE FROM N.A.
RX MEDLINE=95085262; PubMed=7993087;
RA Maciver B., McHale R.H., Saul D.J., Bergquist P.L.;
RT "Cloning and sequencing of a serine proteinase gene from a
RT thermophilic Bacillus species and its expression in Escherichia
RT coli.";
RL Appl. Environ. Microbiol. 60:3981-3988(1994).
RN [2]
X-RAY CRYSTALLOGRAPHY (1.80 ANGSTROMS) OF 122-401.
RX MEDLINE=20057863; PubMed=10588904;
RA Smith C.A., Toogood H.S., Baker H.M., Daniel R.M., Baker E.N.;
RT "Calcium-mediated thermostability in the subtilisin superfamily: the
RT crystal structure of Bacillus Ak.1 protease at 1.8-A resolution.";
RL J. Mol. Biol. 294:1027-1040(1999).
CC -! COFACTOR: Binds 3 calcium ions and 1 sodium ion per subunit.
CC -! SUBCELLULAR LOCATION: Secreted.
CC -! MISCELLANEOUS: Has a pH optimum of 8.5, a temperature optimum of
CC 75 degrees Celsius.
CC -! SIMILARITY: Belongs to peptidase family S8.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; L29506; AAA63688.1; -
CC FIR; I39974; I39974.
CC PDB; 1DBI; 18-NOV-99.
CC MEROPS; S08.009; -.
CC InterPro; IPR000209; Peptidase_S8.
CC InterPro; IPR009020; Protease_inhib.
CC Pfam; PF00082; Peptidase_S8; 1.
CC PRINTS; PR00723; SUBTILISIN.
CC PROSITE; PS00136; SUBTILASE_ASP; 1.
CC PROSITE; PS00137; SUBTILASE_HIS; 1.
CC PROSITE; PS00138; SUBTILASE_SER; 1.
CC Hydrolase; Serine protease; Zymogen; Metal-binding; Calcium-binding;
KW Signal; 3D-structure.
KW SIGNAL; 24
FT SIGNAL 1 24 POTENTIAL.
FT PROPEP 25 121
FT CHAIN 122 401 THERMOPHILIC SERINE PROTEINASE.
FT ACT_SITE 160 160 CHARGE RELAY SYSTEM.
FT ACT_SITE 193 193 CHARGE RELAY SYSTEM.
FT ACT_SITE 347 347 CHARGE RELAY SYSTEM.
FT METAL 126 126 CALCIUM 1.
FT METAL 168 168 CALCIUM 2.
FT METAL 169 169 CALCIUM 3.
FT METAL 171 171 CALCIUM 1.
FT METAL 179 179 CALCIUM 3.
FT METAL 184 184 CALCIUM 2.
FT METAL 186 186 CALCIUM 2.
FT METAL 204 204 CALCIUM 1.
FT METAL 204 204 CALCIUM 3.
FT METAL 207 207 CALCIUM 1.
FT METAL 209 209 CALCIUM 1.
FT METAL 211 211 CALCIUM 1.

```

```

297 297 METAL
300 300 METAL
323 323 METAL
258 258 DISULFID
127 127 TURN
129 129 HELIX
132 132 TURN
135 135 TURN
136 136 TURN
137 137 HELIX
140 140 TURN
140 140 TURN
141 141 HELIX
144 144 TURN
145 145 TURN
147 147 TURN
152 152 TURN
153 153 TURN
155 155 STRAND
166 166 TURN
168 168 TURN
173 173 TURN
174 174 STRAND
179 179 STRAND
183 183 TURN
184 184 STRAND
184 184 STRAND
193 193 HELIX
202 202 TURN
219 219 STRAND
224 224 TURN
228 228 TURN
229 229 HELIX
234 234 HELIX
246 246 TURN
247 247 STRAND
251 251 STRAND
254 254 TURN
263 263 HELIX
274 274 TURN
275 275 TURN
278 278 STRAND
282 282 STRAND
285 285 TURN
298 298 TURN
301 301 STRAND
306 306 TURN
308 308 TURN
312 312 STRAND
315 315 TURN
316 316 STRAND
316 316 TURN
320 320 TURN
324 324 STRAND
327 327 STRAND
331 331 STRAND
333 333 TURN
336 336 TURN
338 338 STRAND
339 339 STRAND
343 343 TURN
346 346 HELIX
362 362 TURN
363 363 TURN
364 364 TURN
376 376 HELIX
377 377 TURN
379 379 STRAND
382 382 TURN
383 383 TURN
384 384 STRAND
385 385 TURN
387 387 TURN
388 388 STRAND
391 391 STRAND
392 392 TURN
395 395 HELIX
399 399 TURN
400 400 TURN
SQ SEQUENCE 401 AA; 42835 MW; 1C736EF4A89F256F CRC64;

Query Match 10.8%; Score 242.5; DB 1; Length 401;
Best Local Similarity 30.9%; Pred. No. 7.1e-09;
Matches 81; Conservative 30; Mismatches 100; Indels 51; Gaps 10;

Qy 11 DVAQSSYGLYQGQGIIVAVADTGLDTGRNDSMHEAPRGKITALLY-ALGRNNANDTNGHG 69
Db 146 DVTKGS-----SGGEIAVIDTGVD-----YTHPDLGKVIKGYDFVDNDYDPMDLNHHG 194
Qy 70 THVAG---SVLGNSTNKGWQANLVFQSIMDSGGGLGSLPSNLQTLFSQAYSAGARTH 126
Db 195 THVAGIAAETNNAITGIAGWPNTRILAVRALDRNG--SGTISDIADAIYADSAREVI 252
Qy 127 TNSWGAAVNGAYTTDSRNVDYVVRKNDMTILFAAGNEGNGGTISAPGAKNAITYGATE 186
Db 253 NLSLGC---DCHTTTLENVAVNWKGSVVVAAAGNNG--SSTTFEPASYENVIAVGA-- 305
Qy 187 NLRPSFGSYADNINHVAAQFSSRGPTKDGRIKPDVMAPGTIFILSARSLAPDSSFWANHDS 246
Db 306 -----VDQYDRLASFSNYGTW-----VDVAPGVDIVSTITG-----N 338

```

QY 247 KYAYMGCTSMATPIVAGNVAQL 268
 Db 339 RYAYMGCTSMASPHVAGLAALL 360

RESULT 12
 SUBS_BAGLE STANDARD; PRT; 269 AA.
 ID SUBS_BAGLE STANDARD; PRT; 269 AA.
 AC P29600;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Subtilisin Savinase (EC 3.4.21.62) (Alkaline protease).
 OS Bacillus lentus.
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 OX NCBI_TaxID=1467;
 RN [1]
 RP X-RAY CRYSTALLOGRAPHY (1.4 ANGSTROMS).
 RX MEDLINE=92148829; PubMed=1738156;
 RA Betzel C., Klupsch S., Papendorf G., Hastrup S., Branner S.,
 RA Wilson K.S.;
 RT "Crystal structure of the alkaline proteinase Savinase from Bacillus
 lentus at 1.4-A resolution.";
 RL J. Mol. Biol. 223:427-445(1992).
 RN [2]
 RP STRUCTURE BY NMR.
 RX MEDLINE=96184541; PubMed=8654411;
 RA Remerowski M.L., Pepermans H.A.M., Hilbers C.W., van de Ven F.J.M.;
 RT "Backbone dynamics of the 269-residue protease Savinase determined
 from 15N-NMR relaxation measurements.";
 RL Eur. J. Biochem. 235:629-640(1996).
 RN [3]
 RP X-RAY CRYSTALLOGRAPHY (0.78 ANGSTROMS).
 RX MEDLINE=98426039; PubMed=9753430;
 RA Kuhn P., Knapp M., Soltis S.M., Ganshaw G., Thoenes M., Bott R.;
 RT "The 0.78-A structure of a serine protease: Bacillus lentus
 subtilisin.";
 RL Biochemistry 37:13446-13452(1998).
 CC -!- FUNCTION: Subtilisin is an extracellular alkaline serine protease,
 it catalyzes the hydrolysis of proteins and peptide amides.
 CC -!- CATALYTIC ACTIVITY: Hydrolysis of proteins with broad specificity
 for peptide bonds, and a preference for a large uncharged residue
 in P1. Hydrolyzes peptide amides.
 CC -!- COFACTOR: Binds 2 calcium ions per subunit.
 CC -!- SUBCELLULAR LOCATION: Secreted.
 CC -!- BIOTECHNOLOGY: Used as a detergent protease. Sold under the name
 Savinase by Novozymes.
 CC -!- MISCELLANEOUS: Secretion of subtilisin is associated with onset of
 sporulation, and many mutations which block sporulation at early
 stages affect expression levels of subtilisin. However, subtilisin
 is not necessary for normal sporulation.
 CC -!- SIMILARITY: Belongs to peptidase family S8.
 DR PDB; 1C9J; 06-OCT-99.
 DR PDB; 1C9M; 10-JAN-01.
 DR PDB; 1C9N; 10-JAN-01.
 DR PDB; 1GCI; 11-NOV-98.
 DR PDB; 1TAV; 11-JUL-01.
 DR PDB; 1JEA; 26-NOV-97.
 DR PDB; 1SVN; 14-OCT-96.
 DR MEROPS; S08.103; --
 DR InterPro; IPR000209; Peptidase S8.
 DR Pfam; PF00082; Peptidase S8; 1.
 DR PRINTS; PR00723; SUBTILISIN.
 DR PROSITE; PS00136; SUBTILASE ASP; 1.
 DR PROSITE; PS00137; SUBTILASE HIS; 1.
 DR PROSITE; PS00138; SUBTILASE SER; 1.
 KW Hydrolase; Sporulation; Serine protease; Metal-binding;
 Calcium-binding; 3D-structure.
 FT ACT_SITE 32 32 CHARGE RELAY SYSTEM.
 FT ACT_SITE 62 62 CHARGE RELAY SYSTEM.
 FT ACT_SITE 215 215 CHARGE RELAY SYSTEM.
 FT METAL 2 2 CALCIUM 1.
 FT METAL 40 40 CALCIUM 1.

FT METAL 73 73 METAL 1 (VIA CARBONYL OXYGEN).
 FT METAL 75 75 CALCIUM 1.
 FT METAL 77 77 CALCIUM 1 (VIA CARBONYL OXYGEN).
 FT METAL 79 79 CALCIUM 1 (VIA CARBONYL OXYGEN).
 FT METAL 163 163 CALCIUM 2 (VIA CARBONYL OXYGEN).
 FT METAL 165 165 CALCIUM 2 (VIA CARBONYL OXYGEN).
 FT METAL 168 168 CALCIUM 2 (VIA CARBONYL OXYGEN).
 FT STRAND 2 2
 FT HELIX 6 10
 FT TURN 11 12
 FT HELIX 13 18
 FT TURN 19 20
 FT TURN 24 25
 FT STRAND 27 32
 FT TURN 39 40
 FT TURN 43 48
 FT TURN 51 52
 FT HELIX 62 71
 FT STRAND 78 78
 FT TURN 84 85
 FT STRAND 87 92
 FT TURN 94 94
 FT TURN 96 97
 FT STRAND 100 100
 FT HELIX 102 114
 FT TURN 115 116
 FT STRAND 119 122
 FT STRAND 126 126
 FT HELIX 131 142
 FT TURN 143 144
 FT STRAND 146 150
 FT STRAND 161 161
 FT TURN 162 164
 FT TURN 166 167
 FT STRAND 168 174
 FT TURN 176 177
 FT STRAND 180 180
 FT TURN 182 183
 FT TURN 188 189
 FT STRAND 190 195
 FT STRAND 199 203
 FT TURN 204 206
 FT STRAND 207 211
 FT HELIX 214 231
 FT TURN 233 234
 FT HELIX 237 246
 FT TURN 247 247
 FT STRAND 249 249
 FT HELIX 254 257
 FT TURN 258 259
 FT STRAND 261 261
 FT HELIX 264 267
 FT TURN 268 269
 SQ SEQUENCE 269 AA; 25698 MW; 4D89F8779999BF8D CRC64;

Query Match 10.7%; Score 240; DB 1; Length 269;
 Best Local Similarity 31.1%; Pred. No. 6.2e-09;
 Matches 91; Conservative 30; Mismatches 90; Indels 82; Gaps 14;

QY 8 VYADVAAQSSYGLYGQGIIVAVADTGLDTGRNDSSNHEAFRGKITALYALGRNNANDTNG 67
 Db 11 VOAPAAHNR-GLTSGVKVAVLDTGIST-----HPLNIRGGASFVPGEP-STQDNG 61
 QY 68 HGTHVAGSV--LNGSTNKGMAPOANLVFQSTIMDSGGGLGGLPSNLQTLFSQAYSAGARI 125
 Db 62 HGTHVAGTIAALNIGVLGVAPSALYAVKVLGASG--SGSVSSIAQGLEWAGNGMHV 119
 QY 126 HTNSWGA-----ANGAVTTSRNVDDYVRKNKNTIILFAAGNEPNGGTISAPGTA 176
 Db 120 ANLSLGSPPSATLEQAVNSA---TSRGV-----LVVAASGNSG--AGSISYPARY 165
 QY 177 KNAITVGTATE--NLRPSFGSYADNINHVAAQSSRGPTKDGRIKPDVMAPGTFILSARSL 234

Db 166 ANAMAVGATDQNNRASFQYAGL-----DIVAPGVNVQSTVPG- 205
QY 235 APDSFWANHDSKYAYMGCTSMATPVAGNVA-----QLRHFVKN 275
Db 206 -----STYASLNGTSMATPVAGAAALVQKNPSNVQIRNH-LKN 246

RESULT 13
ELYA_BACAO
ID ELYA_BACAO STANDARD; PRT; 380 AA.
AC P27693;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Alkaline protease precursor (EC 3.4.21.-).
OS Bacillus alcalophilus.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1445;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PB92;
RX MEDLINE=91282483; PubMed=2059048;
RA van der Laan J.C., Gerritse G., Mulleners L.J.M., van der Hoek R.A.,
Quax W.J.;
RT "Cloning, characterization, and multiple chromosomal integration of a
Bacillus alkaline protease gene.";
RL Appl. Environ. Microbiol. 57:901-909(1991).
RN [2]
RP X-RAY CRYSTALLOGRAPHY (1.75 ANGSTROMS).
RC STRAIN=PB92;
RX MEDLINE=92390330; PubMed=1518788;
RA van der Laan J.C., Teplyakov A.V., Keiders H., Kalk K.H., Misset O.,
Mulleners L.J.M., Dijkstra B.W.;
RT "Crystal structure of the high-alkaline serine protease PB92 from
Bacillus alcalophilus.";
RL Protein Eng. 5:405-411(1992).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (1.85 ANGSTROMS).
RX MEDLINE=93078250; PubMed=1447775;
RA Sobek H., Hecht H.-J., Ashle W., Schomburg D.;
RT "X-ray structure determination and comparison of two crystal forms of
a variant (Asn15Arg) of the alkaline protease from Bacillus
alcalophilus refined at 1.85-A resolution.";
RL J. Mol. Biol. 228:108-117(1992).
RN [4]
RP STRUCTURE BY NMR OF 112-380.
RC STRAIN=PB92;
RX MEDLINE=97277237; PubMed=9115441;
RA Martin J.R., Mulder F.A., Karimi-Nejad Y., van der Zwan J.,
Mariani M., Schipper D., Boelens R.;
RT "The solution structure of serine protease PB92 from Bacillus
alcalophilus presents a rigid fold with a flexible substrate-binding
site.";
RL Structure 5:521-532(1997).
CC -|- COFACTOR: Binds 2 calcium ions per subunit.
CC -|- SUBCELLULAR LOCATION: Secreted.
CC -|- SIMILARITY: Belongs to peptidase family S8.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M65086; AAA2212.1; -;
CC EMBL; A13738; CAA01128.1; -;
CC PIR; A49778; A49778.
CC PDB; 1AH2; 15-APR-98.
CC MEROPS; S08.038; -;
CC InterPro; IPR000209; Peptidase S8.
CC InterPro; IPR009020; Protease_inhib.

DR Pfam: PF00082; Peptidase S8; 1.
DR PRINTS; PR00723; SUBTILISIN.
DR PROSITE; PS00136; SUBTILASE_ASP; 1.
DR PROSITE; PS00137; SUBTILASE_HIS; 1.
DR PROSITE; PS00138; SUBTILASE_SER; 1.
KW Hydrolase; Serine protease; Zymogen; Metal-binding; Calcium-binding;
KW Signal; 3D-structure.
FT SIGNAL 1 27 POTENTIAL.
FT PROPEP 28 112 ALKALINE PROTEASE.
FT CHAIN 113 380 CHARGE RELAY SYSTEM.
FT ACT_SITE 143 173 CHARGE RELAY SYSTEM.
FT ACT_SITE 326 326 CHARGE RELAY SYSTEM.
FT METAL 113 113 CALCIUM 1.
FT METAL 151 151 CALCIUM 1.
FT METAL 184 184 CALCIUM 1 (VIA CARBONYL OXYGEN).
FT METAL 186 186 CALCIUM 1.
FT METAL 188 188 CALCIUM 1 (VIA CARBONYL OXYGEN).
FT METAL 190 190 CALCIUM 1 (VIA CARBONYL OXYGEN).
FT METAL 274 274 CALCIUM 2 (VIA CARBONYL OXYGEN).
FT METAL 276 276 CALCIUM 2 (VIA CARBONYL OXYGEN).
FT METAL 279 279 CALCIUM 2 (VIA CARBONYL OXYGEN).
FT HELIX 117 120
FT TURN 121 123
FT HELIX 125 130
FT TURN 131 131
FT TURN 135 136
FT STRAND 138 142
FT TURN 150 151
FT STRAND 156 157
FT HELIX 173 182
FT STRAND 198 201
FT TURN 207 208
FT HELIX 213 226
FT STRAND 230 233
FT HELIX 242 252
FT TURN 253 255
FT STRAND 257 258
FT TURN 267 268
FT TURN 272 275
FT STRAND 279 279
FT STRAND 282 285
FT STRAND 291 291
FT TURN 299 300
FT STRAND 303 306
FT STRAND 310 314
FT TURN 315 317
FT STRAND 318 322
FT HELIX 325 342
FT TURN 344 345
FT HELIX 348 358
FT HELIX 375 380
SQ SEQUENCE 380 AA; 38853 MW; 539EA72771B6682C CRC64;
Query Match 10.7%; Score 240; DB 1; Length 380;
Best Local Similarity 31.1%; Pred. No. 9.7e-09;
Matches 91; Conservative 30; Mismatches 90; Indels 82; Gaps 14;
QY 8 VKADVAOSSGLYGGQGVAVADTGLDTGRNDSMHEAFRGKITALYALGRTNNDNTNG 67
Db 122 VQAPAAHNR-GLTSGGVKVAVLDTGIST-----HPDLNIRGGASFVPGEP-STQDNG 172
QY 68 HGTHVAGSV--LNGSTNKGNAPOANLVFQSIMDSGGGLGGLPSNLQTLFSAQYAGARI 125
Db 173 HGTHVAGTIAALNNISIGLVGAPNAELYAVKVLGASG--SGSVSSIAQGLEWAGNMGHV 230
QY 126 HTNSNGA-----AVNGAYTDSRVDDYVRKNDMTILFAAGNPGNGGTISAPGTA 176
Db 231 ANLSLGSPPSATLEQAVNSA---TSRGV-----LVVAAGSNGS--AGSISYARY 276
QY 177 KNATITVGATE--NLRPSFGSYADNINHVAFSSRGPTKDKRIKPDVNVAPGTFFILARS 234
Db 277 ANAMAVGATDQNNRASFQYAGL-----DIVAPGVNVQSTVPG- 316


```
QY 235 APDSFWANHDSKYAYMGGTSMATPIVAGNVA-----OLREHFVN 275
Db 317 -----SYASLNGTSMATPHVAGAAALVKQKNSWENVQIRNH-LKN 357

RESULT 14
ELVA_BACCS STANDARD; PRT; 380 AA.
AC P41362;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Alkaline protease precursor (EC 3.4.21.-)
OS Bacillus clausii.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=79880;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=221 / ATCC 21522 / JCM 9139 / DSM 2512;
RX MEDLINE=93043753; PubMed=1368952;
RA Takami H., Kobayashi T., Kobayashi M., Yamamoto M., Nakamura S.,
RA Aono R., Horikoshi K.;
RT "Molecular cloning, nucleotide sequence, and expression of the
RT structural gene for alkaline serine protease from alkaliphilic
RT Bacillus sp. 221.";
RL Biosci. Biotechnol. Biochem. 56:1455-1460 (1992).
RN [2]
RP SEQUENCE OF 112-129.
RC STRAIN=221 / ATCC 21522 / JCM 9139 / DSM 2512;
RA Horikoshi K.;
RL (In) Horikoshi K. (eds.);
RL Microorganisms in alkaline environments, pp.187-194, VCH,
RL Weinheim (1991);
CC -1- COFACTOR: Binds 2 calcium ions per subunit (By similarity).
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: Belongs to peptidase family S8.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; S48754; AAC60420.1; -;
DR EMBL; D13157; BAA02442.1; -;
DR EMBL; A26817; CAA01836.1; -;
DR EMBL; A22550; CAA01611.1; -;
DR HSP; P29600; LC1.
DR MEROPS; S08.103; -;
DR InterPro; IPR000209; Peptidase S8.
DR InterPro; IPR003020; Protease Inhib.
DR Pfam; PF00062; Peptidase_S8; 1.
DR PRINTS; PR00723; SUBTILISIN.
DR PROSITE; PS00136; SUBTILASE ASP; 1.
DR PROSITE; PS00137; SUBTILASE HIS; 1.
DR PROSITE; PS00138; SUBTILASE_SER; 1.
DR KX Hydrolase; Serine protease; Zymogen; Metal-binding; Calcium-binding;
DR Signal.
DR SIGNAL 1 27
DR PROPEP 28 111
DR FT CHAIN 112 380
DR FT ACT_SITE 143 143
DR FT ACT_SITE 173 173
DR FT ACT_SITE 326 326
DR FT METAL 113 113
DR FT METAL 151 151
DR FT METAL 184 184
DR FT METAL 186 186
DR FT METAL 188 188
```

```
FT METAL 190 190 (BY SIMILARITY).
FT FT CALCIUM 1 (VIA CARBONYL OXYGEN)
FT METAL 274 274 (BY SIMILARITY).
FT FT CALCIUM 2 (VIA CARBONYL OXYGEN)
FT METAL 276 276 (BY SIMILARITY).
FT FT CALCIUM 2 (VIA CARBONYL OXYGEN)
FT METAL 279 279 (BY SIMILARITY).
FT FT CALCIUM 2 (VIA CARBONYL OXYGEN)
SQ SEQUENCE 380 AA; 38826 MW; 5F73ABC6D5B6831 CRC64;

Query Match 10.7%; Score 240; DB 1; Length 380;
Best Local Similarity 31.1%; Pred. No. 9.7e-09;
Matches 91; Conservative 30; Mismatches 90; Indels 82; Gaps 14;

QY 8 VKADVAQSSYGLYGGQGIIVADVADTGLDTRNDSSMHEAFKGTALYALQRTNNANTNG 67
Db 122 VQAPAAHNR-GLTSGVKVAVLDGTIST-----HPDLNIRGASFPGEF-STQDNG 172

QY 68 HGTHTVAGSV--LNGSTNKGMAPOANLVFOSIMDSGGGLGLPSNLQTLFSQAYSAGARI 125
Db 173 HGTHTVAGTIAALNNISIGLVGAPSAELYAVKVLGASG--SGSVSSIAQGLEWAGNGMHV 230

QY 126 HTNSWGA-----AVNGAYTTDSRNVDDYVRKNDMTILFAAGNEPNGGTISAPGTA 176
Db 231 ANLSIGSPSPSATLEQAVNSA---TSRGV-----LVVAASGNSG--AGSISYPARY 276

QY 177 KVATVGAATE--NLRPSFGSYADNINHVAFQSSRGPTKDGRIKPDVAVAGPTFILSARSSL 234
Db 277 ANAMAVGATDQNNRRASFQYAGL-----DIVAPGVNVQSTYFG- 316

QY 235 APDSFWANHDSKYAYMGGTSMATPIVAGNVA-----OLREHFVN 275
Db 317 -----SYASLNGTSMATPHVAGAAALVKQKNSWENVQIRNH-LKN 357

RESULT 15
PRTM_BACSP STANDARD; PRT; 269 AA.
AC Q99405;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE M-protease (EC 3.4.21.-).
OS Bacillus sp. (strain KSM-K16).
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1409;
RN [1]
RP X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS).
RA Yamane T., Kani T., Hatanaka T., Suzuki A., Ashida T.,
RA Kobayashi T., Ito S., Yamashita O.;
RT "Structure of a new alkaline serine protease (M-protease) from
RT Bacillus sp. KSM-K16."
RL Acta Crystallogr. D 51:199-206 (1995).
RN [2]
RP SEQUENCE OF 1-23, AND CHARACTERIZATION.
RX MEDLINE=93558832; PubMed=7632397;
RA Kobayashi T., Hakamada Y., Adachi S., Hitomi J., Yoshimatsu T.,
RA Koike K., Kawai S., Ito S.;
RT "Purification and properties of an alkaline protease from
RT alkalophilic Bacillus sp. KSM-K16."
RL Appl. Microbiol. Biotechnol. 43:473-481 (1995).
CC -1- COFACTOR: Binds 2 calcium ions per subunit.
CC -1- SUBUNIT: Monomer.
CC -1- SIMILARITY: Belongs to peptidase family S8.
CC PDB; 1MPT; 22-JUN-94.
DR MEROPS; S08.010; -;
DR InterPro; IPR000209; Peptidase S8.
DR Pfam; PF00082; Peptidase_S8; 1.
DR PRINTS; PR00723; SUBTILISIN.
DR PROSITE; PS00136; SUBTILASE ASP; 1.
DR PROSITE; PS00137; SUBTILASE HIS; 1.
DR PROSITE; PS00138; SUBTILASE_SER; 1.
DR SIGNAL 1 27
DR PROPEP 28 111
DR FT CHAIN 112 380
DR FT ACT_SITE 143 143
DR FT ACT_SITE 173 173
DR FT ACT_SITE 326 326
DR FT METAL 113 113
DR FT METAL 151 151
DR FT METAL 184 184
DR FT METAL 186 186
DR FT METAL 188 188
```

KW Hydrolase; Serine protease; Metal-binding; Calcium-binding;
KW 3D-structure.
FT ACT_SITE 32 32 CHARGE RELAY SYSTEM.
FT ACT_SITE 62 62 CHARGE RELAY SYSTEM.
FT ACT_SITE 215 215 CHARGE RELAY SYSTEM.
FT METAL 2 2 CALCIUM 1.
FT METAL 40 40 CALCIUM 1.
FT METAL 73 73 CALCIUM 1 (VIA CARBONYL OXYGEN).
FT METAL 75 75 CALCIUM 1.
FT METAL 77 77 CALCIUM 1 (VIA CARBONYL OXYGEN).
FT METAL 79 79 CALCIUM 1 (VIA CARBONYL OXYGEN).
FT METAL 163 163 CALCIUM 2 (VIA CARBONYL OXYGEN).
FT METAL 165 165 CALCIUM 2 (VIA CARBONYL OXYGEN).
FT METAL 168 168 CALCIUM 2 (VIA CARBONYL OXYGEN).
FT HELIX 6 10
FT TURN 11 12
FT TURN 13 18
FT TURN 19 20
FT TURN 24 25
FT TURN 27 32
FT TURN 39 40
FT STRAND 43 48
FT TURN 51 52
FT TURN 62 71
FT TURN 84 85
FT STRAND 87 92
FT TURN 96 97
FT HELIX 102 114
FT TURN 115 116
FT STRAND 119 122
FT STRAND 126 126
FT HELIX 131 142
FT TURN 143 144
FT STRAND 146 150
FT STRAND 161 161
FT TURN 162 164
FT TURN 166 167
FT STRAND 169 174
FT TURN 176 177
FT STRAND 180 180
FT TURN 182 183
FT TURN 188 189
FT STRAND 192 195
FT STRAND 199 203
FT TURN 204 206
FT STRAND 207 211
FT TURN 214 231
FT TURN 233 234
FT TURN 237 246
FT TURN 247 247
FT STRAND 249 249
FT HELIX 254 257
FT TURN 258 259
FT STRAND 261 261
FT HELIX 264 267
FT TURN 268 269
SQ SEQUENCE 269 AA; 26723 MW; 7A03C86D534A1D07 CRC64;

Query Match 10.6%; Score 239; DB 1; Length 269;
Best Local Similarity 31.1%; Pred. No. 7.2e-09;
Matches 91; Conservative 30; Mismatches 90; Indels 82; Gaps 14;
QY 8 VKADVAGSSYGLYQGQGVAVADTGLDTGRNDSMHEAFRGKITALYALGRTNANDTNG 67
DB 11 VQAPAHNR-GLTSGVKVAVLDTGIST-----HPDLNIRGGASFVFGEP-STQDNG 61
QY 68 HGTHVAGSV--LQNGSTNKGMFQANLVFQSIQSGGLGLPSNLQTLFQAYAGARI 125
DB 62 HGTHVAGTIAALNNSIGVLGVAPSAELIYAVKVLGASG--SGSVSSIAQGLEWAGNMGHV 119
QY 126 HTNSWGA-----AVNGAYTDSRNVDDYVRKNDMTILFAAGNEGPNGTISAPGTA 176
DB 120 ANLSGSPSPSATLEQAVNSA---TSRGV-----LVVAASGNSG--AGSISYPARY 165

QY 177 KNAITVIGATE--NLRPSFGSYADNINHVAFQSSRGFTKDGRIKPDVMAPGTILSARSSL 234
DB 166 ANAMAVGATDQNNNRASPSQYGAGL-----DIVAPGVNVQSTYFG- 205
QY 235 APDSSEFWANHDSKYAYMGGTSMATPIVAGNVA-----QLRHFVKN 275
DB 206 -----STYASLNCTSMATPHVAGVAALVKQKPSWNSNVOIRNH-LKN 246

Search completed: March 10, 2004, 14:53:39
Job time : 18 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 10, 2004, 14:50:43 ; Search time 45 Seconds
(without alignments)
3042.998 Million cell updates/sec

Title: US-09-985-689A-1-COPY
Perfect score: 2247
Sequence: 1 NDVARGIVKADVAQSSVGLY.....EVQAYNVPVGPQTFFSLAIVN 434

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SPTREMBL_25:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mmc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archheap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|--------|-------------|--------|--------|--------------------|
| 1 | 2242 | 99.8 | 640 | Q3UV9 | Q3uv9 bacillus sp |
| 2 | 2178 | 96.9 | 639 | Q3UV3 | Q3uv3 bacillus sp |
| 3 | 2138 | 95.1 | 434 | Q3AQ0 | Q3aq0 bacillus sp |
| 4 | 1993.5 | 88.7 | 433 | Q3AQ1 | Q3aq1 bacillus sp |
| 5 | 1989.5 | 88.5 | 433 | Q3AQ4 | Q3aq4 bacillus sp |
| 6 | 1982.5 | 88.2 | 433 | Q3AQ2 | Q3aq2 bacillus sp |
| 7 | 506.5 | 22.5 | 1825 | Q3T91 | Q3t91 dictyosteli |
| 8 | 442 | 19.7 | 1702 | Q3T91 | Q3t91 dictyosteli |
| 9 | 409 | 18.2 | 654 | Q3UC9 | Q3uc9 pyrococcus |
| 10 | 401.5 | 17.9 | 561 | Q3UB2 | Q3ub2 thermoaer |
| 11 | 399 | 17.8 | 1239 | Q3FB24 | Q3fb24 streptomyce |
| 12 | 381 | 17.0 | 430 | Q8ENV1 | Q8env1 oceanobacil |
| 13 | 371 | 16.5 | 1253 | Q9FC06 | Q9fc06 streptomyce |
| 14 | 366 | 16.3 | 1102 | P95684 | P95684 streptomyce |
| 15 | 363.5 | 16.2 | 1208 | Q82B14 | Q82b14 streptomyce |
| 16 | 349 | 15.5 | 1245 | Q9R154 | Q9r154 streptomyce |

| | | | | | | |
|----|-------|------|------|----|--------|--------------------|
| 17 | 348 | 15.5 | 1105 | 2 | Q8KH6 | Q8kh6 streptomyce |
| 18 | 348 | 15.5 | 1139 | 16 | Q82I39 | Q82i39 streptomyce |
| 19 | 347.5 | 15.5 | 444 | 16 | Q9KJ7 | Q9kb7 bacillus ha |
| 20 | 338 | 15.0 | 412 | 2 | Q9AER6 | Q9aer6 thermoaner |
| 21 | 337 | 15.0 | 412 | 16 | Q8RC68 | Q8rc68 thermoaner |
| 22 | 336 | 15.0 | 1237 | 2 | Q8GGT4 | Q8ggt4 streptomyce |
| 23 | 329.5 | 14.7 | 824 | 2 | Q45464 | Q45464 bacillus sp |
| 24 | 329.5 | 14.7 | 1220 | 16 | Q9LOA0 | Q9loa0 streptomyce |
| 25 | 327.5 | 14.6 | 435 | 16 | Q8EMJ3 | Q8emj3 oceanobacil |
| 26 | 321.5 | 14.3 | 891 | 1 | Q3635 | Q3635 thermococu |
| 27 | 320.5 | 14.3 | 442 | 16 | Q3I788 | Q3i788 bacillus su |
| 28 | 308.5 | 13.7 | 431 | 2 | Q9S3L6 | Q9s3l6 bacillus sp |
| 29 | 307.5 | 13.7 | 1398 | 1 | Q9P9L1 | Q9p9l1 pyrococcus |
| 30 | 294.5 | 13.1 | 434 | 2 | Q54327 | Q54327 bacillus sp |
| 31 | 292 | 13.0 | 1135 | 1 | Q9P9D1 | Q9p9d1 uncultured |
| 32 | 282.5 | 12.6 | 586 | 16 | Q8PAL8 | Q8pal8 xanthomonas |
| 33 | 281 | 12.5 | 419 | 2 | Q45681 | Q45681 bacillus su |
| 34 | 280.5 | 12.5 | 959 | 16 | Q8PM57 | Q8pm57 xanthomonas |
| 35 | 280 | 12.5 | 799 | 16 | Q9KEM1 | Q9kem1 bacillus ha |
| 36 | 279.5 | 12.4 | 1345 | 1 | Q54437 | Q54437 staphylothe |
| 37 | 275.5 | 12.3 | 621 | 2 | Q9F486 | Q9f486 alteromonas |
| 38 | 275.5 | 12.3 | 621 | 2 | Q53401 | Q53401 alteromonas |
| 39 | 275 | 12.2 | 1101 | 16 | Q82CF0 | Q82cf0 streptomyce |
| 40 | 272.5 | 12.1 | 568 | 16 | Q8PMC0 | Q8pmc0 xanthomonas |
| 41 | 268.5 | 11.9 | 403 | 2 | Q45463 | Q45463 bacillus sp |
| 42 | 268.5 | 11.9 | 814 | 16 | Q82VB3 | Q82vb3 nitrosomona |
| 43 | 267.5 | 11.9 | 575 | 16 | Q8PNW1 | Q8pnw1 xanthomonas |
| 44 | 266 | 11.8 | 1098 | 16 | Q9LI28 | Q9li28 streptomyce |
| 45 | 265.5 | 11.8 | 966 | 16 | Q8PB28 | Q8pb28 xanthomonas |

ALIGNMENTS

RESULT 1

Q3UV9 PRELIMINARY; PRT; 640 AA.
AC Q3UV9
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)
DT 01-JUN-2003 (Tremblrel. 24, Last annotation update)
DE Protease.
GN PROF.
OS Bacillus sp. KSM-KP43.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=109322;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=KP43;
RA Itoh S., Saeki K.;
RT "new protease";
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB051423; BAB55674.2; -
DR GO; GO:0004289; F:subtilase activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR000209; Peptidase_S8.
DR InterPro; IPR007280; PPC.
DR Pfam; PF00082; Peptidase_S8; 1.
DR Pfam; PF04151; PPC; 1.
DR PRINTS; PR00723; SUBTILISIN
DR PROSITE; PS00137; SUBTILASE_HIS; 1.
DR PROSITE; PS00138; SUBTILASE_SER; 1.
SQ SEQUENCE 640 AA; 67991 MW; 4BBAF77E9D592C15 CRC64;

Query Match 99.8%; Score 2242; DB 2; Length 640;
Best Local Similarity 99.8%; Pred. No. 6.9e-122;
Matches 433; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 NDVARGIVKADVAQSSVGLYGGQIVAVADTGLDTRNDSSMHEAFRGKITALYALGRTN 60
Db 207 NDVARGIVKADVAQSSVGLYGGQIVAVADTGLDTRNDSSMHEAFRGKITALYALGRTN 266
QY 61 NANDTINGHGTAVAGSVLGNSTKMGAPQANILVFQISIMDSGGGLGLPSNLQTLFQSAYS 120

```

Db 267 N A N D T N G H G T H V A G S V L G N G S T N K G M A P O A N L V F Q S I M D S S G G L G L P S N L Q T L F S Q A F S 326
QY 121 A G A R I H T N S G A A V N G A Y T T D S R N V D D Y V R K N D M T I L F A A G N E G P N G G T I S A P G T A K N A I 180
Db 327 A G A R I H T N S G A A V N G A Y T T D S R N V D D Y V R K N D M T I L F A A G N E G P N G G T I S A P G T A K N A I 386
QY 181 T V G A T E N L R P S F G S Y A D N I N H V A Q F S S R G P T K D G R I K P D V M A P G T F I L S A R S S L A P D S S F 240
Db 387 T V G A T E N L R P S F G S Y A D N I N H V A Q F S S R G P T K D G R I K P D V M A P G T F I L S A R S S L A P D S S F 446
QY 241 W A N H D S K Y A Y M G G T S W A T P I V A G N V A Q L R E H F V K N R G I T P K P S L L K A A L I A G A A D I G L G Y 300
Db 447 W A N H D S K Y A Y M G G T S W A T P I V A G N V A Q L R E H F V K N R G I T P K P S L L K A A L I A G A A D I G L G Y 506
QY 301 P N G N Q G W R V T L D K S L N V A Y V N E S S L S T S Q K A T Y S F T A T A G K P L K I S I L V M S D A P A S T T A 360
Db 507 P N G N Q G W R V T L D K S L N V A Y V N E S S L S T S Q K A T Y S F T A T A G K P L K I S I L V M S D A P A S T T A 566
QY 361 S V T L V N D L N L V I T A P N G T Q V V G N D F T S P Y N D W D G R N N V E N F I N A P O S G T Y T I E V Q A Y N 420
Db 567 S V T L V N D L N L V I T A P N G T Q V V G N D F T S P Y N D W D G R N N V E N F I N A P O S G T Y T I E V Q A Y N 626
QY 421 V P V G P Q T F S L A I V N 434
Db 627 V P V G P Q T F S L A I V N 640

```

RESULT 2

```

Q9AQR3 Q9AQR3 PRELIMINARY; PRT; 639 AA.
AC Q9AQR3;
DT 01-JUN-2001 (T-EMBLrel. 17, Created)
DT 01-OCT-2002 (T-EMBLrel. 22, Last sequence update)
DT 01-JUN-2003 (T-EMBLrel. 24, Last annotation update)
DE Protease.
GN PROA.
OS Bacillus sp. 9860.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=133778;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=9860;
RX MEDLINE=20568675; PubMed=11118284;
RA Saeki K., Okuda M., Hatada Y., Kobayashi T., Ito S., Takami H.,
RA Horikoshi K.
RT "Novel oxidatively stable subtilisin-like serine proteases from
RT alkaliphilic Bacillus spp.: enzymatic properties, sequences, and
RT evolutionary relationships."
RL Biochem. Biophys. Res. Commun. 279:313-319(2000).
DR EMBL; AB046403; BAB21266.2; -.
DR HSSP; P00782; 1SUP.
DR GO; GO:0004289; F:peptidase activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR002029; Peptidase_S8.
DR InterPro; IPR007280; PPC.
DR Pfam; PF00082; Peptidase_S8; 1.
DR Pfam; PF04151; PPC; 1.
DR PRINTS; PR00723; SUBTILISIN.
DR PROSITE; PS00137; SUBTILASE_HIS; 1.
DR PROSITE; PS00138; SUBTILASE_SER; 1.
SQ SEQUENCE 639 AA; 68185 MW; 316AF6FDBE4FF54 CRC64;

```

Query Match 96.9%; Score 2178; DB 2; Length 639;

Best Local Similarity 96.1%; Pred. No. 3.5e-118;

Matches 417; Conservative 14; Mismatches 0; Indels 0; Gaps 0;

QY 1 NDVARGIVKADVAQSSVGLYGQGIIVAVADTGLDGTGRNDSMHEAFRGKITALYALGRTN 60

Db 206 NDVARGIVKADVAQSSVGLYGQGIIVAVADTGLDGTGRNDSMHEAFRGKITALYALGRTN 265

```

QY 61 N A N D T N G H G T H V A G S V L G N G S T N K G M A P O A N L V F Q S I M D S S G G L G L P S N L Q T L F S Q A F S 120

```

```

Db 266 N A N D T N G H G T H V A G S V L G N G S T N K G M A P O A N L V F Q S I M D S S G G L G L P S N L Q T L F S Q A F S 325
QY 121 A G A R I H T N S G A A V N G A Y T T D S R N V D D Y V R K N D M T I L F A A G N E G P N G G T I S A P G T A K N A I 180
Db 326 A G A R I H T N S G A A V N G A Y T T D S R N V D D Y V R K N D M T I L F A A G N E G P N G G T I S A P G T A K N A I 386
QY 181 T V G A T E N L R P S F G S Y A D N I N H V A Q F S S R G P T K D G R I K P D V M A P G T F I L S A R S S L A P D S S F 240
Db 386 T V G A T E N L R P S F G S Y A D N I N H V A Q F S S R G P T K D G R I K P D V M A P G T F I L S A R S S L A P D S S F 445
QY 241 W A N H D S K Y A Y M G G T S W A T P I V A G N V A Q L R E H F V K N R G I T P K P S L L K A A L I A G A A D I G L G Y 300
Db 446 W A N H D S K Y A Y M G G T S W A T P I V A G N V A Q L R E H F V K N R G I T P K P S L L K A A L I A G A A D I G L G Y 505
QY 301 P N G N Q G W R V T L D K S L N V A Y V N E S S L S T S Q K A T Y S F T A T A G K P L K I S I L V M S D A P A S T T A 360
Db 506 P N G N Q G W R V T L D K S L N V A Y V N E S S L S T S Q K A T Y T T A T A G K P L K I S I L V M S D A P A S T T A 565
QY 361 S V T L V N D L N L V I T A P N G T Q V V G N D F T S P Y N D W D G R N N V E N F I N A P O S G T Y T I E V Q A Y N 420
Db 566 S V T L V N D L N L V I T A P N G T Q V V G N D F T S P Y N D W D G R N N V E N F I N A P O S G T Y T I E V Q A Y N 625
QY 421 V P V G P Q T F S L A I V N 434
Db 626 V P V G P Q T F S L A I V N 639

```

RESULT 3

```

Q9AQR0 Q9AQR0 PRELIMINARY; PRT; 434 AA.
AC Q9AQR0;
DT 01-JUN-2001 (T-EMBLrel. 17, Created)
DT 01-JUN-2001 (T-EMBLrel. 17, Last sequence update)
DT 01-JUN-2003 (T-EMBLrel. 24, Last annotation update)
DE Protease (Fragment).
GN PROB.
OS Bacillus sp. NV1.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=133781;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NV1;
RX MEDLINE=20568675; PubMed=11118284;
RA Saeki K., Okuda M., Hatada Y., Kobayashi T., Ito S., Takami H.,
RA Horikoshi K.
RT "Novel oxidatively stable subtilisin-like serine proteases from
RT alkaliphilic Bacillus spp.: enzymatic properties, sequences, and
RT evolutionary relationships."
RL Biochem. Biophys. Res. Commun. 279:313-319(2000).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8.
DR EMBL; AB046406; BAB21269.1; -.
DR HSSP; P00782; 1SUP.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0004289; F:subtilase activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR002029; Peptidase_S8.
DR InterPro; IPR007280; PPC.
DR Pfam; PF00082; Peptidase_S8; 1.
DR Pfam; PF04151; PPC; 1.
DR PRINTS; PR00723; SUBTILISIN.
DR PROSITE; PS00137; SUBTILASE_HIS; 1.
DR PROSITE; PS00138; SUBTILASE_SER; 1.
KW Hydrolase; Protease; Serine protease.
FT NON_TER 1
FT NON_TER 434
SQ SEQUENCE 434 AA; 45294 MW; 83517EDDB74125D2 CRC64;

```

Query Match 95.1%; Score 2138; DB 2; Length 434;

Best Local Similarity 93.3%; Pred. No. 4.4e-116;

Matches 405; Conservative 20; Mismatches 9; Indels 0; Gaps 0;

```

QY 1 NDVARGIVKADVAQSSVGLYGQGIIVAVADTGLDGTGRNDSMHEAFRGKITALYALGRTN 60

```

```

Db 1 NDVARGIVKADVAQSSYGLYGQGVAVADTGLDTRNDSSMHEAPRGKITAIYALGRTN 60
QY 61 NANDTNHGTHVAGSVLNGSTNKGMAPOANLVFQSIMDSSGGLGGLPSNLQTLFSQAYS 120
Db 61 NANDPNHGTHVAGSVLNGSTNKGMAPOANLVFQSIMDSSGGLGGLPSNLQTLFSQAYS 120
QY 121 AGARIHTNSWGAVNGAYTTDSRNVDDYVRKNDMTILFAAGNEGPNNGGTISAPGTAKNAI 180
Db 121 AGARIHTNSWGAVNGAYTTDSRNVDDYVRKNDMTILFAAGNEGPNNGGTISAPGTAKNAI 180
QY 181 TVGATENLRPSFGSYADNINHVAQFSSRGPTKDGRIKPDVMAFGTIFILSARSSLPDSSF 240
Db 181 TVGATENLRPSFGSYADNINHVAQFSSRGPTKDGRIKPDVMAFGTIFILSARSSLPDSSF 240
QY 241 WANHDSKYAYMGTSMATPIVAGNVAQLREHFVNKRGITPKPSLLKAALIAAGADIGLY 300
Db 241 WANHDSKYAYMGTSMATPIVAGNVAQLREHFVNKRGITPKPSLLKAALIAAGADIGLY 300
QY 301 PNGNQGWGRVTLDKSLNVAVYVNSSLSSTSQKATYSFTATAGKPLKISLVMSDAPASTTA 360
Db 301 PNGNQGWGRVTLDKSLNVAVYVNSSLSSTSQKATYSFTATAGKPLKISLVMSDAPASTTA 360
QY 361 SVTLVNDLNLVITAPNGTVGVNDFTSPYNDNWDGRNVNENVFINAPQSGTITIEVOAYN 420
Db 361 SVTLVNDLNLVITAPNGTVGVNDFTSPYNDNWDGRNVNENVFINAPQSGTITIEVOAYN 420
QY 421 VPVGPQTFFSLAIVN 434
Db 421 VPVGPQTFFSLAIVN 434

RESULT 4
Q9AQR1 PRELIMINARY; PRT; 433 AA.
ID AC Q9AQR1
AC Q9AQR1; 2001 (T-EMBLrel. 17, Created)
DT 01-JUN-2001 (T-EMBLrel. 17, Last sequence update)
DT 01-JUN-2003 (T-EMBLrel. 24, Last annotation update)
DE Protease (Fragment).
GN PROD.
OS Bacillus sp. SD521.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=133780;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=SD521;
RC MEDLINE=20568675; PubMed=11118284;
RA Saeki K., Okuda M., Hatada Y., Kobayashi T., Ito S., Takami H.,
RA Horikoshi K.;
RT "Novel oxidatively stable subtilisin-like serine proteases from
RT alkaliphilic Bacillus spp.: enzymatic properties, sequences, and
RL Biochem. Biophys. Res. Commun. 279:313-319(2000).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8.
DR EMBL; AB046405; BAB21268.1; -
DR HSSP; Q45670; 1DBI.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0004289; F:subtilase activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR000209; Peptidase_S8.
DR InterPro; IPR007280; PPC.
DR Pfam; PF04151; PPC; 1.
DR PRINTS; PR00723; SUBTILISIN.
DR PROSITE; PS00137; SUBTILASE HIS; 1.
DR PROSITE; PS00138; SUBTILASE_SER; 1.
KW Hydrolase; Protease; Serine protease.
FT NON_TER 1
FT NON_TER 433
SQ SEQUENCE 433 AA; 45576 MW; 98A2DF18FE66DDC CRC64;

```

Query Match 88.7%; Score 1993.5; DB 2; Length 433;
 Best Local Similarity 87.6%; Pred. No. 1e-107;

```

Matches 380; Conservative 29; Mismatches 24; Indels 1; Gaps 1;
QY 1 NDVARGIVKADVAQSSYGLYGQGVAVADTGLDTRNDSSMHEAPRGKITAIYALGRTN 60
Db 1 NDVARGIVKADVAQSSYGLYGQGVAVADTGLDTRNDSSMHEAPRGKITAIYALGRTN 60
QY 61 NANDTNHGTHVAGSVLNGSTNKGMAPOANLVFQSIMDSSGGLGGLPSNLQTLFSQAYS 120
Db 61 NANDPNHGTHVAGSVLNGSTNKGMAPOANLVFQSIMDSSGGLGGLPSNLQTLFSQAWN 119
QY 121 AGARIHTNSWGAVNGAYTTDSRNVDDYVRKNDMTILFAAGNEGPNNGGTISAPGTAKNAI 180
Db 121 AGARIHTNSWGAVNGAYTTDSRNVDDYVRKNDMTILFAAGNEGPNNGGTISAPGTAKNAI 180
QY 181 TVGATENLRPSFGSYADNINHVAQFSSRGPTKDGRIKPDVMAFGTIFILSARSSLPDSSF 240
Db 181 TVGATENLRPSFGSYADNINHVAQFSSRGPTKDGRIKPDVMAFGTIFILSARSSLPDSSF 240
QY 241 WANHDSKYAYMGTSMATPIVAGNVAQLREHFVNKRGITPKPSLLKAALIAAGADIGLY 300
Db 241 WANHDSKYAYMGTSMATPIVAGNVAQLREHFVNKRGITPKPSLLKAALIAAGADIGLY 300
QY 301 PNGNQGWGRVTLDKSLNVAVYVNSSLSSTSQKATYSFTATAGKPLKISLVMSDAPASTTA 360
Db 301 PNGNQGWGRVTLDKSLNVAVYVNSSLSSTSQKATYSFTATAGKPLKISLVMSDAPASTTA 360
QY 361 SVTLVNDLNLVITAPNGTVGVNDFTSPYNDNWDGRNVNENVFINAPQSGTITIEVOAYN 420
Db 361 SVTLVNDLNLVITAPNGTVGVNDFTSPYNDNWDGRNVNENVFINAPQSGTITIEVOAYN 419
QY 421 VPVGPQTFFSLAIVN 434
Db 421 VPVGPQTFFSLAIVH 433

RESULT 5
Q9AQR4 PRELIMINARY; PRT; 433 AA.
ID AC Q9AQR4
AC Q9AQR4; 2001 (T-EMBLrel. 17, Created)
DT 01-JUN-2001 (T-EMBLrel. 17, Last sequence update)
DT 01-JUN-2003 (T-EMBLrel. 24, Last annotation update)
DE Protease (Fragment).
GN PROA.
OS Bacillus sp. D6.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=127889;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=D6;
RC MEDLINE=20568675; PubMed=11118284;
RA Saeki K., Okuda M., Hatada Y., Kobayashi T., Ito S., Takami H.,
RA Horikoshi K.;
RT "Novel oxidatively stable subtilisin-like serine proteases from
RT alkaliphilic Bacillus spp.: enzymatic properties, sequences, and
RL Biochem. Biophys. Res. Commun. 279:313-319(2000).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8.
DR EMBL; AB046402; BAB21265.1; -
DR HSSP; Q45670; 1DBI.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0004289; F:subtilase activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR000209; Peptidase_S8.
DR InterPro; IPR007280; PPC.
DR Pfam; PF04151; PPC; 1.
DR PRINTS; PR00723; SUBTILISIN.
DR PROSITE; PS00137; SUBTILASE HIS; 1.
DR PROSITE; PS00138; SUBTILASE_SER; 1.
KW Hydrolase; Protease; Serine protease.
FT NON_TER 1
FT NON_TER 433
SQ SEQUENCE 433 AA; 45576 MW; 98A2DF18FE66DDC CRC64;

```

SQ SEQUENCE 433 AA; 45636 MW; 52087E0A2516107F CRC64;
 Query Match 88.5%; Score 1989.5; DB 2; Length 433;
 Best Local Similarity 87.3%; Pred. No. 1.7e-107;
 Matches 379; Conservative 29; Mismatches 25; Indels 1; Gaps 1;

QY 1 NDVARGIVKADVAQSSYGLYGQGIIVAVADTGLDGRNDSSMHEAFRGKITALVALGRTN 60
 DB 1 NDVARGIVKADVQNNYGLYGQGVVAVADTGLDGRNDSSMHEAFRGKITALVALGRTN 60

QY 61 NNDNTNGHGHVAGSVLGNSTNKGMAPOANLVFQSIMDSGGGLGGLPSNLTLPFOAYS 120
 DB 61 NNDPNGHGHVAGSVLGN-ALNKGMAPOANLVFQSIMDSGGGLGGLPSNLTLPFOAWN 119

QY 121 AGARIHTNSGAAVNGAYTTDSRNVDYVRKNDMTILFAAGNEGPNNGGTISAPGTAKNAI 180
 DB 120 AGARIHTNSGAPVNGAYTANSRQVDEYVRNDMTILFAAGNEGPNNGGTISAPGTAKNAI 179

QY 181 TVGATENLRPSFGSYADNINHVAFSSRGPTKDGRIKPDVMAPGTIFILSARSSLPDSSF 240
 DB 180 TVGATENYRPSFGSIADNPNHIAQFSRGATRDGRIKPDVTAAGTIFILSARSSLPDSSF 239

QY 241 WANHDSKYAYMGTSMTATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAGAADIGLY 300
 DB 240 WANYNSKYAYMGTSMTATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAGATDVGLGY 299

QY 301 PNGNQGWGRVTLDKSLNVAYNVNESSLSSTQKATYSFTATAGPKLKISLWSDAPASTTA 360
 DB 300 PEGDQGWGRVTLDKSLNVAYNVNEATLTGQKATYSFTQAGPKLKISLWTDAPGSTTA 359

QY 361 SVTLVNDLNLVITAPNGTOYVGNDFTPSYDNDWGNVNNVFNAPQSGTITIEVOAYN 420
 DB 360 SVTLVNDLNLVITAPNGQKIVGNDFPSYPYDNDWGNVNNVFNAPQSGTITIEVOAYN 419

QY 421 VPVGPQTFSLAIVN 434
 DB 420 VPSGPQRFSLAIVH 433

RESULT 6
 Q9AQR2
 ID Q9AQR2 PRELIMINARY; PRT; 433 AA.
 AC Q9AQR2;
 DT 01-JUN-2001 (T-EMBLrel. 17, Created)
 DT 01-JUN-2001 (T-EMBLrel. 17, Last sequence update)
 DT 01-JUN-2003 (T-EMBLrel. 24, Last annotation update)
 DE Protease (Fragment).
 GN PROC.
 OS Bacillus sp. Y.
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 OX NCBI_TaxID=133779;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Y;
 RX MEDLINE=20568675; PubMed=1118284;
 RA Saeki K., Okuda M., Hatada Y., Kobayashi T., Ito S., Takami H.,
 RA Horikoshi K.;
 RT "Novel oxidatively stable subtilisin-like serine proteases from
 RT alkaliphilic Bacillus sp.; enzymatic properties, sequences, and
 RT evolutionary relationships.";
 RL Biochem. Biophys. Res. Commun. 279:313-319(2000).
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8.
 DR EMBL; AB046404; BAB21267.1; -;
 DR HSSP; Q45670; 1DBI.
 DR GO; GO:0008233; F:peptidase activity; IEA.
 DR GO; GO:0004289; F:subtilase activity; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR000209; Peptidase_S8.
 DR InterPro; IPR007280; PPC.
 DR Pfam; PF00082; Peptidase_S8; 1.
 DR Pfam; PF04151; PPC; 1.
 DR PRINTS; PR00723; SUBTILISIN.
 DR PROSITE; PS00137; SUBTILASE_HIS; 1.

DR PROSITE; PS00138; SUBTILASE_SER; 1.
 KW Hydrolase; Protease; Serine protease.
 FT NON_TER 1
 FT NON_TER 433 433
 SQ SEQUENCE 433 AA; 45587 MW; B81291A803C775AE CRC64;
 Query Match 88.2%; Score 1982.5; DB 2; Length 433;
 Best Local Similarity 87.1%; Pred. No. 4.4e-107;
 Matches 378; Conservative 30; Mismatches 25; Indels 1; Gaps 1;

QY 1 NDVARGIVKADVAQSSYGLYGQGIIVAVADTGLDGRNDSSMHEAFRGKITALVALGRTN 60
 DB 1 NDVARGIVKADVQNNYGLYGQGVVAVADTGLDGRNDSSMHEAFRGKITALVALGRTN 60

QY 61 NNDNTNGHGHVAGSVLGNSTNKGMAPOANLVFQSIMDSGGGLGGLPSNLTLPFOAYS 120
 DB 61 NNDPNGHGHVAGSVLGN-ALNKGMAPOANLVFQSIMDSGGGLGGLPSNLTLPFOAWN 119

QY 121 AGARIHTNSGAAVNGAYTTDSRNVDYVRKNDMTILFAAGNEGPNNGGTISAPGTAKNAI 180
 DB 120 AGARIHTNSGAPVNGAYTANSRQVDEYVRNDMTILFAAGNEGPNNGGTISAPGTAKNAI 179

QY 181 TVGATENLRPSFGSYADNINHVAFSSRGPTKDGRIKPDVMAPGTIFILSARSSLPDSSF 240
 DB 180 TVGATENYRPSFGSIADNPNHIAQFSRGATRDGRIKPDVTAAGTIFILSARSSLPDSSF 239

QY 241 WANHDSKYAYMGTSMTATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAGAADIGLY 300
 DB 240 WANYNSKYAYMGTSMTATPIVAGNVAQLREHFVKNRGITPKPSLLKAALIAGATDVGLGY 299

QY 301 PNGNQGWGRVTLDKSLNVAYNVNESSLSSTQKATYSFTATAGPKLKISLWSDAPASTTA 360
 DB 300 PEGDQGWGRVTLDKSLNVAYNVNEATLTGQKATYSFTQAGPKLKISLWTDAPGSTTA 359

QY 361 SVTLVNDLNLVITAPNGTOYVGNDFTPSYDNDWGNVNNVFNAPQSGTITIEVOAYN 420
 DB 360 SVTLVNDLNLVITAPNGQKIVGNDFPSYPYDNDWGNVNNVFNAPQSGTITIEVOAYN 419

QY 421 VPVGPQTFSLAIVN 434
 DB 420 VPSGPQRFSLAIVH 433

RESULT 7
 Q8T9W1
 ID Q8T9W1 PRELIMINARY; PRT; 1825 AA.
 AC Q8T9W1;
 DT 01-JUN-2002 (T-EMBLrel. 21, Created)
 DT 01-JUN-2002 (T-EMBLrel. 21, Last sequence update)
 DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
 DE Serine protease/ABC transporter TagD.
 GN TAGD.
 OS Dictyostelium discoideum (Slime mold).
 OC Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.
 OX NCBI_TaxID=44689;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=AX4;
 RA Ajlard C., Loomis W.F.;
 RT "Evolution of the ABC transporters of Dictyostelium.";
 RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.
 DR EMBL; AF466309; AAL74253.1; -;
 DR GO; GO:0016020; C:membrane; IEA.
 DR GO; GO:0005524; F:ATP binding; IEA.
 DR GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti...; IEA.
 DR GO; GO:0000166; F:nucleoside binding; IEA.
 DR GO; GO:0008233; F:peptidase activity; IEA.
 DR GO; GO:0004289; F:subtilase activity; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR GO; GO:0006810; P:transport; IEA.
 DR InterPro; IPR003593; AAA ATPase.
 DR InterPro; IPR001140; ABC_TM_transpt.

```
DR InterPro; IPR003439; ABC transporter.
DR InterPro; IPR000209; Peptidase S8.
DR Pfam; PF00664; ABC membrane; 1.
DR Pfam; PF00005; ABC trans; 1.
DR Pfam; PF00082; Peptidase S8; 1.
DR PRINTS; PR00723; SUBTILISIN.
DR ProDom; PD000006; ABC transporter; 1.
DR SMART; SM00382; AAA; 1.
DR PROSITE; PS00211; ABC TRANSPORTER_1; 1.
DR PROSITE; PS00893; ABC_TRANSPORTER_2; 1.
DR PROSITE; PS00137; SUBTILASE HIS; 1.
DR PROSITE; PS00136; SUBTILASE SER; 1.
KW ATP-binding; Protease; Transport.
SQ SEQUENCE 1825 AA; 202641 MW; E28160BC78613A3B CRC64;

Query Match      22.5%; Score 506.5; DB 5; Length 1825;
Best Local Similarity 27.7%; Pred. No. 9e-21;
Matches 163; Conservative 77; Mismatches 155; Indels 193; Gaps 22;

QY 19 LYGGQGVAVADTGLDTR---NDS-----SMHEAFRGKITALYALGRTNNANDTNGH 68
DB 327 LRKGQQLSADTGLDGHCHFFSDSNPNIPYNSVNLNHRXVTVYIGSL--HDNEDYVDGH 384
QY 69 GTHVAGSVLNG-----STNKGAPOANLVFQSI-MDSGGGLGLPLSNLQTLFQOAYS 120
DB 385 GTHVCGSAAGAPEDSSLAISFSLATDAKIAFDLASPSNNEPVPEDYSQLYOFLYN 444
QY 121 AGARIHNSGA-----AVNGAYTDSRDVDYVRKN-DMTILFAAGNEGNGGTISAPGT 175
DB 445 AGARVHGDGSLSIQGLYASYSDDAGSIDDFLYTHPDFILRAAGNNEQYSLLS-QAT 503
QY 176 AKNAITVGATENLRPSF-----
DB 504 AKNIVTGAETHTHESITTDALYSNFTVAKSTLNSLQSPDKDKYCTYTTAQCCTEYST 563
QY 193 -----GSYAD-----NINHVAQFSRGPTKDGRIKPDVMAFGTIFLSA 230
DB 564 VKLSGCGCTSYIKNSYASIFSSQPELYNENNICSFSSKGPHTDGLKPDIVAPQGVITSA 623
QY 231 RSSLA-----PDSFVWHDISKYAMGTSNATPIVAGNVAQLREH----- 271
DB 624 RSNQANTDQCGDGLSNLTALLSE-----SGTSMATPLATAANTILQYLVGDYPT 676
QY 272 --FYKRGITPKPRLKAAALIAGAADIGLGY-----NGNQMGWV 310
DB 677 GSIVESNKLQPTGSLKALINNAQLNGTFPLSNTNPSNAVDFTFAGANFVQWGS 736
QY 311 TLDKSLNVAVYNESS-----SLTSOKATYST----- 338
DB 737 RMSEWL---YVESGVRKPSRWVIGELGKDKKASNNKEYSLSLQGVNSYCTYKPS 793
QY 339 --ATAGKP-LKISLVWSDAPASTTASVTLVNDLNLVITAPNGTQ---YVGNDFTSYND 391
DB 794 GSNSGGIPRIVATLVWTDPSYSGAKLNLVNDLNTW---NTESEFIFYNSGGSSYNG 850
QY 392 N-----WDGRNNVNF---INAPSGTYTIEVQAVVPGPQTF 429
DB 851 TKGTTPLQDSINNVEGIIVTPINTKGEISFRFIAGTNIPIGPQNF 898

RESULT 8
Q9GTN7 PRELIMINARY; PRT; 1702 AA.
AC Q9GTN7;
DC 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE TAGA.
GN TAGA.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
CX NCBI_TaxID=44689;
RN [1]
```

```
RP SEQUENCE FROM N.A.
RA Good J.R., Cabral M., Kuspa A.;
RT "TagA, a putative serine protease/ABC transporter of Dictyostelium
RT that is expressed at the onset of development and is required for the
RT differentiation of a subpopulation of prespore cells.";
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY.
DR EMBL; AF263455; AAG11416.1; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0004009; F:ATP-binding cassette (ABC) transporter acti...; IEA.
DR GO; GO:0000166; F:nucleotide binding; IEA.
DR GO; GO:0004289; F:subtilase activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR003593; AAA ATPase.
DR InterPro; IPR001140; ABC TM transpt.
DR InterPro; IPR003439; ABC transporter.
DR InterPro; IPR00209; Peptidase S8.
DR Pfam; PF00664; ABC membrane; 1.
DR Pfam; PF00005; ABC trans; 1.
DR Pfam; PF00082; Peptidase S8; 1.
DR PRINTS; PR00723; SUBTILISIN.
DR ProDom; PD000006; ABC transporter; 1.
DR SMART; SM00382; AAA; 1.
DR PROSITE; PS00211; ABC TRANSPORTER_1; 1.
DR PROSITE; PS00893; ABC_TRANSPORTER_2; 1.
KW ATP-binding; Transport.
SQ SEQUENCE 1702 AA; 187103 MW; 4A67716303CB7131 CRC64;

Query Match      19.7%; Score 442; DB 5; Length 1702;
Best Local Similarity 25.6%; Pred. No. 4.4e-17;
Matches 139; Conservative 100; Mismatches 164; Indels 140; Gaps 22;

QY 18 GYGGQGVAVADTGLDTR---NDSSMHEAFRGKITALYALGRTNNANDTNGHGHVAG 74
DB 300 GIKGGEIVGCADTGIDINHCFFYDTNPIGSTHRIIS-YSSGNGDOTDEIDGHTIVG 358
QY 75 SVLNGNSTN-----KGMAPQANLVFQSI-MDSGGGLGLPLSNLQTLFQOAYSAGARIHT 127
DB 359 TIIGTTVDPSVSEFSGGAPNSKVAFDVLQVSGNGL-SIQSNLTAIYQSYDQNAKVC 417
QY 128 NSGAAVNGAYTDSRDVDYVRKN-DMTILFAAGNEGNGG-----TISAPGTAKNAIT 181
DB 418 DAWNSNIGFFYTGVTGTEMIDRFQWDHPDPLVYVRSAGN-NVNFNFSIVTLTSQESTSKNSLV 476
QY 182 VCATENLRPSGVSADNINH-----
DB 477 VGSNN-QPS-STYLLSIDYDWDPIYNSIRTSVCTQCSIYGTCDSPVPTQTSVDIQT 533
QY 202 -----VAQFSRGPTKDGRIKPDVMAFGTIFLS 229
DB 534 QCCSNPILAKICCSPEIQOQYQTNSTVYSEFIPSLFSGVGTSDGRLKPDLLAPGSPIS 593
QY 230 ABSSLAPDSSFWANH-----DSKYAMGTSNATPIVAGNVAQLRE-----HFVK 275
DB 594 SR-SLGPSTI--NHCSPTISGATISALIAMEGSSQAAVATSAALVLRQYRDGYFING 650
QY 276 R-----GITPKPSLLKAAALIAGAA---DIGLYPNNGQWGRVTLDSKLNVAVYNESS-- 325
DB 651 KYNSSVGFQPSASLVKATLINTASINVDTLEY---SQPGNIQLSKLITTTNAQTSLD 707
QY 326 -----SLTSOKATYSTATAGKPLKISLVWSDAPASTTASVTLVNDLNLVITA- 374
DB 708 IPSSTIEKADPIINTGETNSYCFSLDSKADIDITLVWTDPAQSPLSTFTLVNLDLALLAF 767
QY 375 PNG--TOYVGNDFTSYNDN---WDGRNNVNFINAPSGTYTIEVQAVVPGPQTF 429
DB 768 VDGELSIYSGNSEITFKNTSQVIFDQNLNNEVIRIKDAPIGSYDVKIFGTNIVPNQSY 827
QY 430 LAI 432
DB 828 VVI 830
```

```
RESULT 9
Q8UB09 PRELIMINARY; PRT; 654 AA.
AC Q8UB09;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Alkaline serine protease.
PF1670.
OS Pyrococcus furiosus.
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC Pyrococcus.
OX NCBI_TaxID=2261;
[1]
RN SEQUENCE FROM N.A.
RC STRAIN=Vcl / DSM 3638 / ATCC 43587 / JCM 8422;
RA Weiss R.B., Dunn D.M., Robb F.T., Brown J.R.;
RT "The complete sequence of the Pyrococcus furiosus genome.";
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; A5010265; AL81794.1; -
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0004289; F:subtilase activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR000209; Peptidase_S8.
DR Pfam; PF00082; Peptidase_S8; 1.
DR PRINTS; PR00723; SUBTILISIN.
DR PROSITE; PS00136; SUBTILASE ASP; 1.
DR PROSITE; PS00137; SUBTILASE HIS; 1.
DR PROSITE; PS00138; SUBTILASE_SER; 1.
KW Protease; Complete proteome.
SQ SEQUENCE 654 AA; 70230 MW; 1CB145A5F505DB34 CRC64;

Query Match 18.2%; Score 409; DB 17; Length 654;
Best Local Similarity 29.6%; Pred. No. 1e-15;
Matches 135; Conservative 60; Mismatches 147; Indels 114; Gaps 19;

QY 18 GYGQGGQIVAVADTGLDTRGNDSSMHEAFRGKITALVALGRTNAN-----DTNGHGH 71
DB 154 GYDGGGITTIGDITGID-----ASHFDLQGGKV-----IGWDFVNGRSYPYDDHGHGH 202
QY 72 VAGSVLGNSTN-----KGMAPQANLVFQSIM--DSGGGLGGI.PSNLQTLFSAQYASAGARI 125
DB 203 VASIAAGTGAASNGKYKGAPAGAKLAGIKVLGADGGSGSTIIKGVWAVDNKKYGIKV 262
QY 126 HTNSWA-----AVNGAYTDSRNVDYVRKNDMTILFAAGNEGPGGTISAP 173
DB 263 INLSGSSQSSDGTALSOAVNAAWDA-----GLVVVVAAGNSGNGKTYIGSP 310
QY 174 GTAKNATVGTENLRPSFGSYADNINHYAQSSRGPTKDGRIKPDVWAPGTFFLSABSS 233
DB 311 AAASKVITVGA-----VDKYDVITSFSSRGPTADGRLKPEVWAPGNWIIAARAS 359
QY 234 LAPDSSFANHDSKYAYMGTSMATPIVAGNVAQLREHFVKRGITPK--PSLLKAALIA 291
DB 360 ---GTSMGQPINDYYTAAPGTSMATPHVAGIAALLIQ-----AHPSTPDKVKTALIE 409
QY 292 GA-----ADTGLGYPNGQGWRTLDKSLNVAYNSSLSSTQKA-----TYSFT 338
DB 410 TADIVKPEADIATAYGA-----GRVNAYKAIN--YDNYAKLVFTGYVANKSGSTHQFV 460
QY 339 ATAGKPLKISLVNSDAPASTASVTLVNDNLNIVITAPNGTQVVGNDFTSPYNDNDGRNN 398
DB 461 ISGASFVATLYNDNAN-----SDLDLYLDPNGNQ-VDYSYIAY-----G 501
QY 399 VENVFNAPSGTYTIEVQAVNVVPGQPTFSIAVN 434
DB 502 FEKVGYNPTDGTWTKVWSYS---GSANYQVDWS 534

RESULT 10
Q8RBJ2 PRELIMINARY; PRT; 561 AA.
AC Q8RBJ2;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Subtilisin-like serine proteases.
APR2 OR TTE0824.
OS Thermoanaerobacter tengcongensis.
OC Bacteria; Firmicutes; Clostridia; Thermoanaerobacteriales;
OC Thermoanaerobacteriaceae; Thermoanaerobacter.
OX NCBI_TaxID=119072;
[1]
RN SEQUENCE FROM N.A.
RC STRAIN=MB4 / JCM 11007;
MEDLINE=21992816; PubMed=11997336;
RA Bao Q., Tian Y., Li W., Xu Z., Xuan Z., Hu S., Dong W., Yang J.,
RA Chen Y., Xue Y., Xu Y., Lai X., Huang L., Dong X., Ma Y., Ling L.,
RA Tan H., Chen R., Wang J., Yu J., Yang H.;
RT "A complete sequence of T. tengcongensis genome.";
RL Genome Res. 12:689-700(2002).
DR EMBL; AE013049; AAM24081.1; -
DR GO; GO:0005509; F:calcium ion binding; IEA.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0004289; F:subtilase activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR000209; Peptidase_S8.
DR InterPro; IPR007280; PPC.
DR InterPro; IPR009020; Protease_inhib.
DR Pfam; PF00082; Peptidase_S8; 1.
DR Pfam; PF04151; PPC; 1.
DR PRINTS; PR00723; SUBTILISIN.
DR PROSITE; PS00136; SUBTILASE ASP; 1.
DR PROSITE; PS00137; SUBTILASE HIS; 1.
DR PROSITE; PS00138; SUBTILASE_SER; 1.
KW Protease; Complete proteome.
SQ SEQUENCE 561 AA; 59968 MW; BA9C5C52F7083A18 CRC64;

Query Match 17.9%; Score 401.5; DB 16; Length 561;
Best Local Similarity 31.4%; Pred. No. 2.3e-15;
Matches 143; Conservative 63; Mismatches 135; Indels 119; Gaps 22;

QY 6 GIVKADVAQSSYGLYGQGIIVAVADTGLDTRGNDSSMHEAFRGKITALVALGRTNAN-- 63
DB 155 GIVK---ARSDFGVTGKNTIIAIDIGDNHVDLS-----GGKI-----IGWKDFINK 201
QY 64 ---DTNGHGHVAGSVLGNSTN---KGMAPQANLVFQSIMDSGGG-----LGLPSN 110
DB 202 TTPYDDNGHGHVAVASIAAGTGAGNSFYKGVAPDALLVGIKVLGDANGSGSMSTVTAGIDWA 261
QY 111 LQ-----TLPQSAQYAGARIHTNSWGAANGAYTTDSRNVDYVRKNDMTILF 158
DB 262 VQNKVYGVKIVNLSLGTSTSSDG---TDTSLVN-----RAVD-----SGIVVV 305
QY 159 AAGNEGPGGTISAPGTAKNAITVGTENLRPSFGSYADNINHYAQSSRGPTKDGRIK 218
DB 306 AAGNSGPAKTYTIGSPGAERKAITVAAMADV---GELGFNL---ASFSSRGPTADGRIK 358
QY 219 DWAPGTFFLSABSSAPDSSFWANHDSKYAYMGTSMATPIVAGNVAQLREHFVKRGI 278
DB 359 DIAPAGYNIITAAK-----ANSVNGYVYTSYSGTSMATPFVAGTVALMLN---ANPNL 405
QY 279 TPKPSLLKAALIAAGADIGLGYPNQGWGRVRLDKSLNVA-----YNES 324
DB 406 TPDA--KNIMSTAKSWGPPSKNDYAGRLQGYEARVAGFRGNIDVPNHYYI--S 461
QY 325 SSLSTQKA--TYSFTAT-AGKPLKISLV---WSDAPASTASVTLVNDNLNIVITAPNGTQ 379
DB 462 GYLPGRSYSDTWTFTNATNTSYPTAITLIIIPDWANYP-----DFDIYLDPSGTL 511
QY 380 YVGNDFTSYNDNDWGRNNVENVFNAPSGTYTIEVQAY 419
```



```

Db 173 EYDD-NHGTGTHCAGDAAGNALSQYQGPAPDANLVGVKVLNKTGS-GSLSTVIEGI- 229
QY 116 SOAYSAGARIHNSWGAANGAVYTTDSRVVDYVR-----KNDMTILFAAGNEGPGGT 169
Db 230 DMCIQNQSYKNINILSL-GSDATEPASGDPVNAVETAWDNGVMVCAAGNSGPGDKT 288
QY 170 ISAPGTAKNAITVGATENLRPFSGYADNHNHVAQFSSRPGTKQRIKPDVNAVPTFILS 229
Db 289 VSGPSGIPKVIITVGAADDNNTAERS-----DDSAEFSSRPGTIDGLTKPNLLTPGVDIVS 344
QY 230 ARS--SLAPDSFMANHDSKYAYMGTSMTATPVAQNLREHFVKNRGITPKPSLLKA 287
Db 345 LRAPGSFIDKTNKSARVGSNYLSGTSMTATPICIGIVAQLLQ---SDSLT--PNQVKE 399
QY 288 ALIAGAADTGLGVPNGQNGRVTLDKSLNVAVNE 323
Db 400 KLMEACQDLGQS-PN-VQAGYLV---NAANLININE 430

RESULT 13
Q9FC06 PRELIMINARY; PRT; 1253 AA.
AC Q9FC06;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Putative secreted peptidase.
GN SC07176 OR SC8A11.04C.
OS Streptomyces coelicolor.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Saunders D.C., Harris D.;
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Cerdeno A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RX MEDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kieser H.M., Denapante D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96(1996).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2) / M145;
RX MEDLINE=21996410; PubMed=12000953;
RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA Cronin A., Frazer A., Goble A., Hdaigo J., Hornsby T., Howarth S.,
RA Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neill S.,
RA Rabinowitch E., Rajandream M.A., Rutherford K., Rutter S.,
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA Hopwood D.A.;
RT "Complete genome sequence of the model actinomycete Streptomyces
RT coelicolor A3(2).";
RL Nature 417:141-147(2002).
RL EMBL; AL939130; CAC01576.1; -.
DR HSSP; Q99405; IMPT.
DR GO; GO:0008233; F:peptidase activity; IEA.
DR GO; GO:0004289; F:subtilase activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.

```

```

DR InterPro; IPR00137; PA.
DR InterPro; IPR000209; Peptidase_s8.
DR Pfam; PF02225; PA; 1.
DR PRINTS; PR00723; SUBTILISIN.
DR PROSITE; PS00136; SUBTILASE ASP; 1.
DR PROSITE; PS00137; SUBTILASE HIS; 1.
DR PROSITE; PS00138; SUBTILASE SER; 1.
KW Complete proteome.
SQ SEQUENCE 1253 AA; 130971 MW; AA69B417EFEEB89 CRC64;

Query Match 16.5%; Score 371; DB 16; Length 1253;
Best Local Similarity 27.4%; Pred. No. 3 9e-13;
Matches 141; Conservative 59; Mismatches 145; Indels 170; Gaps 22;

QY 16 SYGLYGQGIIVAVADTGLDTGRNDSMHEAFRKITATYALGRNTNANTNGHTGHVAGS 75
Db 233 SGGNTGEGVGVAVLDTGVDAG-----HFDFAAGRIATAATASFVPDQDVTDRNGHTHVA 286
QY 76 VLNGSTN---KGMAPQANLVFQSIMDSGG-----GLGG 106
Db 287 VAGTGAASGGVEKGVAPGASLHGKVKVLDNSGQSDSWVLAGWEAVRDQHAKIVNSLGD 346
QY 107 LPSNQLTFSQAYSAGARIHNSWGAANGAVYTTDSRVVDYVRKNDMTILFAAGNEGPN 166
Db 347 SPTDGTDLPLSEAV-----NWLSAETGA-----LFVVAAGNSGPE 380
QY 167 GGTISAPGTAKNAITVGATENLRPFSGYADNHNHVAQFSSRGP-TKDGRIKPDVNAVPT 225
Db 381 AYTGTGPAADAALTGVANG--PGKG-----VDQLADFSSRGPVGDNAVRDLTPAGV 433
QY 226 FILSARSLAPDSSFANHDSKYAYMGTSMTATPVAQNLREHFVKNRGITPKPSLL 285
Db 434 GVLAARSRYAPEG-----EGAYQSLSGTSMATPHVAGAAALLAAEHPDWTG-----QRL 482
QY 286 KALLIAGAADIGLYPENGQ-----GNGRVTLDKSLNVAVYNESSLSLTSQ---KATY 335
Db 483 KEALVGTTA-----GTQRFSPFDAGSGRV-----DVAAAVRSLTLAGSDAFQAHY 528
QY 336 -----SFTATAGKPLKISLVWSDA-----PASTTASVTLVND 367
Db 529 PYTPGQTVRRDVTYNTSGPAFVALDLALSPAELPEGLFTLSEAQVTVPAHGTSVGVITH 588
QY 368 LN-----LVITAPNGTQYVNDFTSPYNDWDGRNVDNENVFVNA-----PQSG 410
Db 589 LDAEDNGAYATLVASGADGAVLA----RTPVGVNKEGRR--ATLALTAKDHHDKPLSG 642
QY 411 T-----YTIEVQA-YNVVGVGPTFSL 430
Db 543 TVILKQVETNATPKVYSDASGELDRLSPSTYSV 677

RESULT 14
P95684 PRELIMINARY; PRT; 1102 AA.
AC P95684;
DT 01-MAY-1997 (TrEMBLrel. 03, Created)
DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Subtilisin-like protease.
OS Streptomyces albobacillus.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1887;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S-3253;
RX MEDLINE=97144528; PubMed=8990295;
RA Suzuki M., Taguchi S., Yamada S., Kojima S., Miura K., Momose H.;
RT "A novel member of the subtilisin-like protease family from
RT Streptomyces albobacillus.";
RL J. Bacteriol. 179:430-438(1997).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8.

```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: March 15, 2004, 19:18:46 ; Search time 3605 Seconds
(without alignments)

5217.996 Million cell updates/sec

Title: US-09-985-689a-1-COPY

Perfect score: 2247
Sequence: 1 NDVARGIVKADVAQSSVGLY.....EVQAYNPVPGPTFFSLAIVN 434

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Deiop 6.0 , Delext 7.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlh
-Q=/csm2_1/USPTO_epool/SWOP8985689/runat_10032004_112806_19360/app_query.fasta_1.583
-DB=GenEmbl -QMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=SWOP8985689 @CGN 1 1 2372 @runat_10032004_112806_19360 -NCFU=6 -ICPU=3
-NO_WMAP -LARGESQUERY -NEG_SCORES=0 -runat_10032004_112806_19360 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPEXT=10 -XGAPEXT=0.5 -FGAPOPOP=6
-FGAPEXT=7 -YGAPOPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl:*
1: gb_ba:*
2: gb_hcg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_mu:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*

29: em_vi:*
30: em_hcg_hum:*
31: em_hcg_inv:*
32: em_hcg_other:*
33: em_hcg_mus:*
34: em_hcg_pln:*
35: em_hcg_rod:*
36: em_hcg_mam:*
37: em_hcg_vrt:*
38: em_sy:*
39: em_hcg_hum:*
40: em_hcg_mus:*
41: em_hcg_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Match | Length | DB | ID | Description |
|------------|--------|-------|--------|----|-----------|----------------------|
| 1 | 2242 | 99.8 | 1305 | 6 | AX839476 | Sequence |
| 2 | 2242 | 99.8 | 1923 | 1 | AB051423 | AB051423 Bacillus |
| 3 | 2242 | 99.8 | 1923 | 6 | AB368117 | AB368117 Sequence |
| 4 | 2237 | 99.6 | 1923 | 1 | AB084155 | AB084155 Bacillus |
| 5 | 2237 | 99.6 | 1923 | 6 | AB368118 | AB368118 Sequence |
| 6 | 2178 | 96.9 | 1920 | 1 | AB046403 | AB046403 Bacillus |
| 7 | 2178 | 96.9 | 1920 | 6 | AB368116 | AB368116 Sequence |
| 8 | 2138 | 95.1 | 1302 | 1 | AB046406 | AB046406 Bacillus |
| 9 | 2120.5 | 94.4 | 3003 | 6 | AR069954 | AR069954 Sequence |
| 10 | 2120.5 | 94.4 | 3003 | 6 | BD062155 | BD062155 Nucleic a |
| 11 | 1993.5 | 88.7 | 1299 | 1 | AB046405 | AB046405 Bacillus |
| 12 | 1989.5 | 88.5 | 1299 | 1 | AB046402 | AB046402 Bacillus |
| 13 | 1982.5 | 88.2 | 1299 | 1 | AB046404 | AB046404 Bacillus |
| 14 | 1981.5 | 88.2 | 2218 | 6 | EO3808 | EO3808 DNA encodin |
| C 15 | 627.5 | 27.9 | 60006 | 1 | AF268611 | AF268611 Unculture |
| 16 | 585 | 26.0 | 301205 | 1 | AF017218 | AF017218 Geobacter |
| 17 | 519.5 | 23.1 | 5890 | 3 | DU600086 | DU600086 Dictyosteli |
| 18 | 506.5 | 22.5 | 5772 | 3 | AF466309 | AF466309 Dictyosteli |
| 19 | 492 | 21.9 | 6115 | 3 | DU204432 | DU204432 Dictyosteli |
| 20 | 447.5 | 19.9 | 1977 | 6 | AR201152 | AR201152 Sequence |
| 21 | 442 | 19.7 | 5109 | 3 | AF263455 | AF263455 Dictyostate |
| C 22 | 420.5 | 18.7 | 126928 | 2 | AC096673 | AC096673 Trypanoso |
| 23 | 411.5 | 18.3 | 1236 | 6 | AR201146 | AR201146 Sequence |
| 24 | 411.5 | 18.3 | 1962 | 6 | AR201155 | AR201155 Sequence |
| C 25 | 411.5 | 18.3 | 12452 | 1 | AR010265 | AR010265 Pyrococcu |
| C 26 | 401.5 | 17.9 | 14376 | 1 | AB013049 | AB013049 Thermoana |
| C 27 | 399 | 17.8 | 303450 | 1 | SC0939130 | SC0939130 Streptomy |
| C 28 | 381 | 17.0 | 299850 | 1 | AF004601 | AF004601 Oceanobac |
| 29 | 366 | 16.3 | 3348 | 1 | D83672 | D83672 Streptomyce |
| 30 | 363.5 | 16.2 | 300425 | 1 | AP005044 | AP005044 Streptomy |
| 31 | 353 | 15.7 | 2539 | 6 | AR202322 | AR202322 Sequence |
| 32 | 349 | 15.5 | 291000 | 1 | SC0939105 | SC0939105 Streptomy |
| 33 | 348 | 15.5 | 2809 | 6 | AR202321 | AR202321 Sequence |
| 34 | 348 | 15.5 | 3900 | 1 | AB007809 | AB007809 Streptomy |
| C 35 | 348 | 15.5 | 302300 | 1 | AF005034 | AF005034 Streptomy |
| 36 | 347.5 | 15.5 | 292550 | 1 | AP001513 | AP001513 Bacillus |
| 37 | 338 | 15.0 | 1239 | 1 | AF305633 | AF305633 Thermoana |
| 38 | 338 | 15.0 | 1239 | 1 | AY028704 | AY028704 Thermoana |
| C 39 | 337 | 15.0 | 12295 | 1 | AE013026 | AE013026 Thermoana |
| 40 | 336 | 15.0 | 135638 | 1 | AF484556 | AF484556 Streptomy |
| C 41 | 332.5 | 14.8 | 1329 | 6 | AX433519 | AX433519 Sequence |
| C 42 | 331 | 14.7 | 207829 | 1 | BSUB0010 | Z99113 Bacillus su |
| C 43 | 329.5 | 14.7 | 6854 | 1 | BSU39230 | U39230 Bacillus sp |
| C 44 | 329.5 | 14.7 | 300800 | 1 | SC0939112 | AL939112 Streptomy |
| C 45 | 327.5 | 14.6 | 301150 | 1 | AP004602 | AP004602 Oceanobac |

ALIGNMENTS

RESULT 1

AX839476
LOCUS AX839476 1305 bp DNA linear PAT 15-DEC-2003
Sequence 2 from Patent EP1347044.
ACCESSION AX839476
VERSION AX839476.1 GI:39922766
KEYWORDS
SOURCE Bacillus sp. KSM-KP43
ORGANISM Bacillus sp. KSM-KP43
Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
REFERENCE
1 Okuda, M.K., Sato, T.K., Saito, K.K., Sumitomo, N.K., Izawa, Y.K.,
Saeki, K.K., Kobayashi, T.K. and Nomura, M.K.
Alkaline protease
Patent: EP 1347044-A 2 24-SEP-2003;
Kao Corporation (JP)
FEATURES
Location/Qualifiers
1..1305
/organism="Bacillus sp. KSM-KP43"
/mol_type="unassigned DNA"
/db_xref="taxon:109322"
<1..1305
/note="unnamed protein product"
/codon_start=1
/transl_table=11
/db_xref="GI:39922767"
/protein_id="CAE85146.1"
/translation="NDVARGIVKADYAGSSVLYCQGIQVAVADTGLDTRGNDSSMHE
APRKITADYALGRINNDINGHTHVGSLNGSTNGKNGAPQANLVFOSIMDSGG
GUGLPSNLQIFLSQAYAGARIHTNSWGAAYTTDSRVDYVYKNDMTILFAA
GNEPNGTISAPGAKNAITVGTANTENRPSGYADNINHVAFSSRGPDKGRIFK
DYVAPGTFLSARSLADPSFWANHDSKYAVMGSTNATVAGNVAQLRHFVYK
GITPKPSLLKALIAAGADIGLYPNGQWGRVTLDKSLNVAVYVNSSLSLTSOKAT
YSFTATAGKPLAISLVMSDAPASTTASVLTNVDLDLVITAPNGIYVGNDFTSFYNDN
WGRNVNENFINAQSGTYTIEVQAYNVVPGPQFSLAIVN"
ORIGIN
Alignment Scores:
Pred. No.: 1,19e-133 Length: 1305
Score: 2242.00 Matches: 433
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.78% Mismatches: 0
Query Match: 99.78% Indels: 0
DB: 6 Gaps: 0
US-09-985-689A-1-COPY (1-434) x AX839476 (1-1305)
QY 1 AsnAspValAlaArgGlyLeuValValAlaAspValAlaGlnSerSerTyrGlyLeuTyr 20
DB 1 AATGATTTGGCGTGGCAATTTCTCAAGCGGATGTGGCTCAGAGCAGCTACGGTTGTAT 60
QY 21 GlyGlnGlyLeuValAlaValAlaAspThrGlyLeuAspThrGlyValAsnAspSer 40
DB 61 GGACAAAGGACAGATCGTAGCGTTGCCGATACAGGCTTGATACAGGTCCGAATGACAGT 120
QY 41 SerMetHisGluAlaPheArgGlyValThrAlaLeuTyrAlaLeuGlyValArgThrAsn 60
DB 121 TCATCATGATGAGCCCTTCGCGGGGAAATTTCTGCAATATATGCAATGGAGCGGCAAT 180
QY 61 AsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
DB 181 AATGCCAATGATACGAATGGTTCATGTAGCATGTGGCTGGCTCCGTATTAGGAAACGGC 240
QY 81 SerThrAsnIleGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100
DB 241 TCACATTAATAAGGAATCGCGCTCAGCGCAATCTAGTCTTCCATCTATCATGTGGATAGC 300
QY 101 GlyGlyGlyLeuGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyrSer 120
DB 301 GGTGGGGGACTTGGAGACTACCTTCGATCTGCAACCTTATTGAGCCAGCATACAGT 360
QY 121 AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAsnGlnValAlaTyrThr 140
DB 361 GCTGGTGCCAGATTCATACAACTCTCGGGGAGCAGCAGTGAATGGGGCTTACACACA 420

QY 141 AspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAla 160
DB 421 GATTCACGAATGTGGATGACTATGTGCGCAAAATGATATGACGATCTCTTTTGGTGGC 480
QY 161 GlyAsnGluGlyProAsnGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 180
DB 481 GCGAATGAAGGACCGAACCGGAAACCATCAGTGCACCGGACACGCTAAAATGCAATA 540
QY 181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 200
DB 541 ACAGTCGGAGCTACGGAAACCTCCGCCCAAGCTTTGGCTTATCGGACATAATCAAC 600
QY 201 HisValAlaGlnPheSerSerArgGlyProThrLysAspGlyValGlyLeuVal 220
DB 601 CATGTGGCAGATTCTCTTACGTGACCGACCAAGGATGGAGGATCAACCGGATGTC 660
QY 221 MetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 240
DB 661 ATGGCACCGGAAACGTTCTACTATCAGCAAGATCTTCTTGGACCGGATTCCTCTCTC 720
QY 241 TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIle 260
DB 721 TGGCGGAACCATCACAGTAATAATATGATATCATGCGTGGTGAACGCTCCATGCGTAC 780
QY 261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPro 280
DB 781 GTTGTGGAAACGTGGCAGCAGCTTCGTGAGCATTTTGTGAATAACAGAGGCATCACCCA 840
QY 281 LysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspIleGlyLeuGlyTyr 300
DB 841 AAGCTTCTCTATTAATAAGCGGCACCTGATTCGGTGGTGCAGCTGCATCGGCTTGGCTAC 900
QY 301 ProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 320
DB 901 CCGAACGCTAACCAAGATGGGACGAGTGACATTTGATATAATCTCCCTGAACGTTGCTAT 360
QY 321 ValAsnGluSerSerSerLeuSerThrSerGlnLysAlaThrTyrSerPheThrAlaThr 340
DB 961 GTGAACGAGTCCAGTTCTCTATCCACACGCAAAAGACGACGCTACTCTGTTTACTGTACT 1020
QY 341 AlaGlyLysProLeuLysIleSerLeuValTrpSerAspAlaProAlaSerThrAla 360
DB 1021 GCCGGCAAGCCTTTGAAATCTCCCTGGTATGCTGATGGCCCTCGAGCACAACCTGCT 1080
QY 361 SerValThrLeuValAsnAspLeuAsnLeuValIleThrAlaProAsnGlyThrGlnTyr 380
DB 1081 TCCGTAAACGCTTCTCAATGATCTGGACCTTGCTATTCACCTCCAAATGGCACACAGTAT 1140
QY 381 ValGlyAsnAspPheThrSerProTyrAsnAspAsnTrpAspGlyArgAsnValGlu 400
DB 1141 GTAGGAAATGACTTTACTTCGCCATACATGATTAATCTGGGATGGCCGCAATACGTAGAA 1200
QY 401 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 420
DB 1201 AATGATTTATTAATGCACCAAGCGGAGCGTATACAAATTGAGGTACAGGCTTATAAC 1260
QY 421 ValProValGlyProGlnThrPheSerLeuAlaIleValAsn 434
DB 1261 GTACCGGTTGGACACAGACCTTCTCGTTGGCAATTGTGAAT 1302
RESULT 2
AB051423
LOCUS AB051423 1923 bp DNA linear BCT 10-MAY-2002
DEFINITION Bacillus sp. KP43 PROF gene for protease, complete cds.
ACCESSION AB051423
VERSION AB051423.2 GI:20521154
KEYWORDS
SOURCE Bacillus sp. KSM-KP43
ORGANISM Bacillus sp. KSM-KP43
Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
REFERENCE
1 Itoh, S. and Saeki, K.
AUTHORS

new protease
 Unpublished
 2 (bases 1 to 1923)
 Saeki, K.
 Direct Submission
 Submitted (21-NOV-2000) Katsuhisa Saeki, KAO CORPORATION;
 2606, AKABANE, ICHIKAIMACHI, HAGA, TOCHIGI 321-3486, Japan
 (E-mail: 387185@kasanet.kao.co.jp, Tel: 81285687471 (ex. 7471),
 Fax: 81285687403)
 On May 9, 2002 this sequence version replaced gi:14164344.
 Location/Qualifiers
 1..1923
 /organism="Bacillus sp. KSM-KP43"
 /mol_type="genomic DNA"
 /strain="KP43"
 /db_xref="taxon:109322"
 1..1923
 /gene="PROF"
 1..1923
 /gene="PROF"
 /codon_start=1
 /transl_table=11
 /product="protease"
 /protein_id="BAB55674.2"
 /db_xref="GI:20521155"
 /translation="MKKKVFLSVLSAAILSTVALSNPSAGGARNFPLDFKGIQTT
 TDARKSKQSGTGAAFLVSENVKLPKGLQKLETPVANNKLHIIQFNGPILSETKQ
 QLEKTKGILDIYIPDYAVIIVEYEGDVKSATSTIEHVESVEPLPIIYRIDPQILFKAS
 ELVKAVALDTKQKNKEVQLRGIEQIAQFAISNDVLYITAKPEYKVNVDARGIVKADV
 AQSYGLYGGQIVAVADTGLTDRNDSSMHEAFRGKITALYALGRNTNANDTNGHGT
 HVAAGVLSNGSTNKGPAQANLVQSIIMDSGGGLGPNLQTLFQAYSAQARIHTN
 SWGAANVAGTTSRNVDDVVRKNDMTILFAAGNPGNGTISAPCTAKNAITVGATE
 NLPSFGSYADNINHVAFSSRGPTKDGRIKPYVMAPGPFILSARSLAPDSFVANH
 DSKYANGSTMAPIVAGNVQAREHFVKNRGLTFKPSLLKAAALAGAADIGLGPFN
 DGKQWGRVTDLSKNVAYNSESLSSTSKATYSFTATACKPLKISLWSDAPATTA
 SVTLVNDLDLVIYAPNGTQYGVNDFTSPYNDNDWRNVENFINAPQSGTITIEVQA
 YNVPVGPQTFSLAIVN"

ORIGIN
 Alignment Scores:
 Pred. No.: 1.92e-133 Length: 1923
 Score: 2242.00 Matches: 433
 Percent Similarity: 100.00% Conservative: 1
 Best Local Similarity: 99.77% Mismatches: 0
 Query Match: 99.78% Indels: 0
 DB: 1 Gaps: 0
 US-09-985-689A-1-COPY (1-434) x AB051423 (1-1923)

Qy 1 AsnAspValAlaArgGlyLeuValValAlaAspValAlaGlnSerSerTyrGlyLeuTyr 20
 Db 619 AATGATGTCGCGTGGAATTGCAAGCGGATGTGGCTCAGAGCAGCTACGGGTTGTAT 678
 Qy 21 GlyGlnGlyGlnLeuValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
 Db 679 GGACAGGACAGATCGTAGCGGTGGCGATACAGGGCTTGATACAGCTCGCAATGACAGT 738
 Qy 41 SerMethHisGluAlaPheArgGlyLeuThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
 Db 739 TCGATGATGAAGCTTCCGCGGAAATTTACTGCAITATATGCAITGGACGACGAAT 798
 Qy 61 AsnAlaAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
 Db 799 AATGCCAATGATACGAATGTCATGTGTACGATGTGGCTGGCTCCGTTATAGGAACGCG 858
 Qy 81 SerThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100
 Db 859 TCCACTAATAAGAAATGGCGCTCAGCGCAATCTAGTCTTCCAACTATCATGGATAGC 918
 Qy 101 GlyGlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyrSer 120
 Db 919 GGTGGGGGACTTGAGGACTACCTTCGAATCTGCAACCTTATTACGCCAAGCATACAGT 978

121 AlaGlyAlaArgIleHisThrAsnSerTyrGlyAlaAlaValAsnGlyAlaTyrThr 140
 Db 979 GCTGCTGCCGAATTCATACAACTCTCTGGGAGCAGCAGTGAATGGGCTTACACAACA 1038
 Qy 141 AspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAla 160
 Db 1039 GATTCAGAAAATGTGGATGACTATGTGCGCAAAATATATAGCAGATCTTTTCGCTGCC 1098
 Qy 161 GlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 180
 Db 1099 GGGATGAAGGACCGACGCGGACCATCAGTCAGCCAGGACAGCTAAAAATGCAATA 1158
 Qy 181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 200
 Db 1159 ACAGTCGGAGCTACGGAAAACCTCGGCCAAAGCTTTGGTCTTATGCGGACAATATCAAC 1218
 Qy 201 HisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVal 220
 Db 1219 CATGTGGCAGATCTCTTCACGTGGACCGCAAGGATGGACGGATCAACCCGATGTC 1278
 Qy 221 MetAlaProGlyThrPheIleLeuSerAlaArgSerIleAlaProAspSerSerPhe 240
 Db 1279 ATGGCACCAGGACGTTCTACTATCAGAAAGATCTTCTCTTGCACCGGATTCCTCCTC 1338
 Qy 241 TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIle 260
 Db 1339 TGGCGAAACCATGACAGTAAATATGCATACATGGGTGGAACGTCCTCATGGCTACCCGATC 1398
 Qy 261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPro 280
 Db 1399 GTTGTCTGAAACGTTGGCACAGCTTCGTGAGCATTTTGTGAAAAACAGAGGCATCACACA 1458
 Qy 281 LysProSerLeuLeuLysAlaAlaIleAlaGlyAlaAlaAspIleGlyLeuGlyTyr 300
 Db 1459 AAGCCTTCTCTATTAAGACGGGCTGATTCGGGTGAGCTGAGCTGGGCTTGGCTAC 1518
 Qy 301 ProAsnGlyAsnGlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 320
 Db 1519 CGRAACGTTAACCAAGGATGGGACGAGTGACATGGATTAATCCCTGAACGTTGCCTAT 1578
 Qy 321 ValAsnGluSerSerSerLeuSerThrSerGlnLysAlaThrTyrSerPheThrAlaThr 340
 Db 1579 GTGAACGAGTCCAGTTCTTATCCACGACCAAAAGCGACGATCTCGTTTACTGTCTACT 1638
 Qy 341 AlaGlyLysProLeuLysIleSerLeuValTyrSerAspAlaProAlaSerThrAla 360
 Db 1639 GCCGCAAGCCTTTGAAATCTCCCTGGTATGGTCTGATGCCCTGGAGACACACTGCT 1698
 Qy 361 SerValThrLeuValAsnAspLeuAsnLeuValIleThrAlaProAsnGlyThrGlnTyr 380
 Db 1699 TCCGTAACGCTTGTCAATGATCTGGACCTTGTCTATTACCGCTCCAAATGGCACACAGTAT 1758
 Qy 381 ValGlyAsnAspPheThrSerProTyrAsnAspAsnTyrAspGlyArgAsnValGlu 400
 Db 1759 GTAGGAATGACTTACTTCCCATCATCATGATTAACCTGGGATGGCCGCAATACGATGAA 1818
 Qy 401 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 420
 Db 1819 AATGTATTATTATATGACCAACAAAGCGGACGTATACATAGGTTAGGTACAGGCTTATAAC 1878
 Qy 421 ValProValGlyProGlnThrPheSerLeuAlaIleValAsn 434
 Db 1879 GTACCGGTTGGACACACAGACCTTCTCGTTGGCAATTTGGAAT 1920

RESULT 3
 AR368117
 LOCUS
 DEFINITION Sequence 5 from patent US 6376227.
 ACCESSION AR368117
 VERSION AR368117.1 GI:34601778
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.

Unclassified.
1 (bases 1 to 1923)
Takaiwa,M., Okuda,M., Saeki,K., Kubota,H., Hitomi,J., Kegeyama,Y.,
Shikata,S. and Nomura,M.
Alkaline protease
Patent: US 6376227-A 5 23-APR-2002;
Location/Qualifiers
1..1923
/organism="unknown"
/mol_type="genomic DNA"

ORIGIN

Alignment Scores:
Pred. No.: 1,92e-133 Length: 1923
Score: 2442.00 Matches: 433
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.77% Mismatches: 0
Query Match: 99.78% Indels: 0
DB: 6 Gaps: 0

US-09-985-689A-1-COPY (1-434) x AR368117 (1-1923)

Qy 1 AsnAspValAlaArgGlyLeuValAlaAspValAlaGlnSerSerTyrGlyLeuTyr 20
Db 619 AATGATGTTCCGGTGGGAATTTCAAGCGGATGGCTCAGACAGCTACGGGTGTAT 678
Qy 21 GlyGlnGlyLeuValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db 679 GGACAAAGGACAGATCGTAGCGGTTGCCGATACAGGCTGTGATACAGGTCGCAATGACAGT 738
Qy 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
Db 739 TCGATCGATGAAGCCTTCGCGGGGAAATTAATCTGCATATATGCAATTCGGAGCGGAAT 798
Qy 61 AsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
Db 799 AATGCCAATGATACGAATGCTCATGTACGATGTGGCTGGCTCGGTATAGGAAACGGC 858
Qy 81 SerThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100
Db 859 TCCACTAAATAAGGAATGGCGCTCAGCGGAATCTAGTCTTCCATCTATCATGTATGATGC 918
Qy 101 GlyGlyGlyLeuGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyrSer 120
Db 919 GTGTGGGGACTTGGAGGACTACCTTCGATCTGCAACCTTATTCAGCCAGCATACAGT 978
Qy 121 AlaGlyAlaArgIleHisThrAsnSerTyrGlyAlaAlaValAsnGlyAlaTyrThr 140
Db 979 GCTGGTGCCAGAAATTCATAAACTCCTGGGAGCAGCAGTGAATGGGCTTACACAACA 1038
Qy 141 AspSerArgAsnValAlaAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAla 160
Db 1039 GATTCAGAAATGTGATGACTATGTGGCNAATAATGATATACGATCCTTTTCGCTGCC 1098
Qy 161 GlyAsnGlyGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 180
Db 1099 GGGAAATGAAGACCGGACCGGGAACCATCAGTCGACACGAGCAGCAGTAAATAATGCAATA 1158
Qy 181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 200
Db 1159 ACAGTGGAGCTACGGAATAACCTCCGCCAAGCTTTGGGTCTTATTCGGACAATATCAAC 1218
Qy 201 HisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVal 220
Db 1219 CATGTGGCACAGTCTCTTCACGTGGACCGCAAAAGGATGGACGATCAACCCGATGTC 1278
Qy 221 MetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerPhe 240
Db 1279 ATGGACCGGGAACGTTCTACTATCATCAGCAAGATCTTCTTTCGACCGGATTCCTCCCTTC 1338
Qy 241 TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyLysThrSerMetAlaThrProIle 260
Db 1339 TGGCGCAACCATCACAGTAAATATGATACATACATGGGTGGAACTCCATGCTACACCGATC 1398

261 ValAlaGlyAsnValAlaGlnLeuArgGlyHisPheValLysAsnArgGlyIleThrPro 280
1399 GTTGTGGAAACGTGGCACAGCTTCTGTAGCATTTTGTGAAACACAGAGCATCACACCA 1458
Qy 281 LysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspIleGlyLeuGlyTyr 300
Db 1459 AAGCCTTCTCTATTAAAGCGGCACATGATTGCGGTGCAGCTGACATCGGCTTGGCTAC 1518
Qy 301 ProAsnGlyAsnGlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 320
Db 1519 CCGAACGGTAACCAAGATGGGACAGTGCATTCGATAAATCCCTGACCGTTGCTAT 1578
Qy 321 ValAsnGluSerSerSerLeuSerThrSerGlnLysAlaThrTyrSerPheThrAlaThr 340
Db 1579 GTGAACGAGTCCAGTTCCTCTATCCACGACGCAAAAGACGACGTACTCGTTTACTGCTACT 1638
Qy 341 AlaGlyLysProLeuLysIleSerLeuValTyrSerAspAlaProAlaSerThrAla 360
Db 1639 GCCGGCAAGCCTTTGAAATCTCCCTGGTATGGTCTGATGCCCTCGGACGACACTGCT 1698
Qy 361 SerValThrLeuValAsnAspLeuValIleThrAlaProAsnGlyThrGlnTyr 380
Db 1699 TCCGTAAACGTTGTCAATGATCTGGACCTTGTCATTACCGCTCCAAATGGCACACAGTAT 1758
Qy 381 ValGlyAsnAspPheThrSerProTyrAsnAspAsnTyrAspGlyArgAsnValGlu 400
Db 1759 GTAGGAAATCAGTTTACTTCGCATCAATGATACTGGGATGGCGCAATAACGTAAGA 1818
Qy 401 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 420
Db 1819 AATGATTTATTAATGCCACCAAAAGCGGACGCTATCAATGAGGTACAGGCTTATAAC 1878
Qy 421 ValProValGlyProGlnThrPheSerLeuAlaIleValAsn 434
Db 1879 GTACCGGTTGGACACAGACCTTCTCGTTGGCAATTGTGAAT 1920

RESULT 4
AB084155
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source

AB084155
Bacillus sp. KSM-9865 gene for protease, complete cds.
AB084155
AB084155.1 GI:34392386
Bacillus sp. KSM-9865
Bacillus sp. KSM-9865
Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
1
Okuda,M., Saeki,K. and Kobayashi,T.
Bacillus sp. KSM-9865 protease gene
Published Only in Database (2003)
2 (bases 1 to 1923)
Okuda,M., Saeki,K. and Kobayashi,T.
Direct Submission
Submitted (18-APR-2002) Mitsuyoshi Okuda, Kao corporation,
Biological Science Laboratory, Ichikaimachi Akabane 2606, Haga,
Tochigi 321-3497, Japan (E-mail:okuda.mitsuyoshi@kao.co.jp,
Tel:81-285-68-7543, Fax:81-285-68-7547)
Location/Qualifiers
1..1923
/organism="Bacillus sp. KSM-9865"
/mol_type="genomic DNA"
/strain="KSM-9865"
/db_xref="taxon:192495"
1..1923
/codon_start=1
/transl_table=11
/product="protease"
/protein_id="BAC82522.1"
/db_xref="GI:34392387"
/translation="MRKKKVLFLVLSAAAILSTVALSNPSAGARNFOLDPKGIOTI
TDKGSKQQTGAAPLVESENKLPKGLQKLETPANNKLHIQVFNPILEETKQ
QLEKTKAKILDYIPDVAYIVEBGDKVKSATSTIEDVESVEPYLPIVRIDPQLPTKGAS

ELVXAVALDTNKKKEVQLRGIEQIAQFATSDVLYITAKPEYKVMNDVARGIVKADV
 AQSSTGLYGGCGTIVAVADTGLDTRNDSSMHEAPRGKITALYALGRNNDANDNGHT
 HVAGSVLGVNGTGMQAPQANLVFQSDMSGGGLGFLPSNLQTLFQAYAGAGARIHTN
 SWRAVNGAYTTDSRNVDYVVRKNDMTILFAAGNEGPNGGTISAPGTAKNAITVGATE
 NLRSPFGYADINIHVAQFSRGRTPKDRIKPDVMAPEGPTFLSARSLAPDPSFWANH
 DSKVAYMGTSMAPIVAVAGNAQLRHFVKNRGTPKPSLLKALIAAGADIGLVEN
 GNOGWRVTLKSLNVAVNESLSSTOKATYSPTATAGKELKISLWSDAPASPTA
 SVTLVNDLIDLITAPNGTQYVGNDFTSFYNNWDRNZNVENFINAQSGTYTIEVQA
 YNVFPGPTFSLAIVN"

ORIGIN

Alignment Scores:

Pred. No.: 3,99e-133 Length: 1923
 Score: 2237.00 Matches: 432
 Percent Similarity: 100.00% Conservative: 2
 Best Local Similarity: 99.54% Mismatches: 0
 Query Match: 99.55% Indels: 0
 DB: 1 Gaps: 0

US-09-985-689A-1-COPY (1-434) x AB084155 (1-1923)

Qy 1 AsnAspValAlaArgGlyLeuValAlaAspValAlaGlnSerSerTyrGlyLeuTyr 20
 Db 619 AATGATGTTGCACGTGGAAATTTGCAACGGATGTGGCGACAGCAGCTACGGGTTGTAT 678
 Qy 21 GlyGlnGlyGlnIleValAlaAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
 Db 679 GGACAAGACAGATCGTAGCGGTTGCCGATACAGGGCTTGATACAGTCCGATGACAGT 738
 Qy 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
 Db 739 TCGATGTCATGAAGCCTTCCCGGGGAAATTTACTGCATTATATGCAATGGGACGACGAAT 798
 Qy 61 AsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
 Db 799 AATGCCAATGATACGAATGGTCATGGTACGATGTGGCTGGCTCCGTATTAGGAAACGGC 858
 Qy 81 SerThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100
 Db 859 TCCACTAATAAGGAATCGCGCTCAGCGGAATCTAGTCTTCCAAATCTATCATGATGATG 918
 Qy 101 GlyGlyGlyLeuGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyrSer 120
 Db 919 GGTGGGGACTTGGAGGACTACCTTCGNAATCTGCACACCTTATTACGCCAGCATACAGT 978
 Qy 121 AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAsnGlyAlaTyrThr 140
 Db 979 GCTGGTGCAGAAATTCATAAATCTCTGGGAGCAGCAGTGAATGGGCTTACACAACA 1038
 Qy 141 AspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAla 160
 Db 1039 GATTCAGAAATGTGGATGATGTGTGGCAAAAATGATATGACGATCCCTTTTCGCTGCC 1098
 Qy 161 GlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 180
 Db 1099 CGGAATGAAGACCGAAGCGCGAACCATCATGTCACCCAGCAGCAGCTAAATATGCAATA 1158
 Qy 181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 200
 Db 1159 ACAGTCGAGACTACGAAACCTCCGCCCAAGCTTCGGGCTCTTATGGCGCAATATCAAC 1218
 Qy 201 HisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVal 220
 Db 1219 CATGTGGCACAGTCTCTTTCAGTGGACCGCAAGGNTGGACGATCAACACCGATGTC 1278
 Qy 221 MetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 240
 Db 1279 ATGGCACCGGAAACGTTCTACTATCAGCAAGATCTTCTCTTGCACCGGATTCCTCCTTC 1338
 Qy 241 TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIle 260
 Db 1339 TGGGCAACCATGACAGTAATATATGATGATGGTGGACGTCCATGGCTACACCGATC 1398

Qy 261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPro 280
 Db 1399 GTTCTGGAAACGTGGCACAGCTTCGTGAGCATTTTGTGAAAAACAGAGGCATCACACA 1458
 Qy 281 LysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspIleGlyLeuGlyTyr 300
 Db 1459 AAGCCTTCTATTAAAGCGGCACTGATTCGCGGTGGCAGCAGACATCGGCTTGGCTAC 1518
 Qy 301 ProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 320
 Db 1519 CCGAACGGTAACCAAGGATGGGACGAGTGACATTCGATAAATCCCTAAACGTTGCCTAT 1578
 Qy 321 ValAsnGluSerSerSerLeuSerThrSerGlnLysAlaThrTyrSerPheThrAlaThr 340
 Db 1579 GTGACAGCAGTCAGATCTCTATCCACCAGCCAAAAGCGACGCTACTCGTTTACTGCTACT 1638
 Qy 341 AlaGlyLysProLeuLysIleSerLeuValTrpSerAspAlaProAlaSerThrThrAla 360
 Db 1639 GCGCGCAGCCTTTGAAGATCTCCCTGGTATGCTGTGATGCTGCGGACGACAACTGCT 1698
 Qy 361 SerValThrLeuValAsnAspLeuAsnLeuValIleThrAlaProAsnGlyThrGlnTyr 380
 Db 1699 TCCGTAAACGCTTGTCAATGACCTGGACCTTGTCAATCCGCTCCAAATGGCACAAATAT 1758
 Qy 381 ValGlyAsnAspPheThrSerProTyrAsnAspAsnTrpAspGlyArgAsnValGlu 400
 Db 1759 GTTGGAAATGACTTTACTTCGCCATACAAATAATACTGGGATGGCGCAATAACGCTAGAA 1818
 Qy 401 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 420
 Db 1819 AATGATATTATTAAATGCGCCACAAAGCGGAGCGTATACAAATGGAGTACAGGCTTATAAC 1878
 Qy 421 ValProValGlyProGlnThrPheSerLeuAlaIleValAsn 434
 Db 1879 GTACCGGTGGACACAGACCTTCTCGTTGGCAATTGTGAAC 1920

RESULT 5
 AR368118
 LOCUS AR368118
 DEFINITION Sequence 7 from patent US 6376227.
 ACCESSION AR368118
 VERSION AR368118.1 GI:34601779
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.
 REFERENCE 1 (bases 1 to 1923)
 AUTHORS Takaiwa, M., Okuda, M., Saeki, K., Kubota, H., Hitomi, J., Kageyama, Y., Shikata, S. and Nomura, M.
 TITLE Alkaline protease
 JOURNAL Patent: US 6376227-A 7 23-APR-2002;
 FEATURES Location/Qualifiers
 source 1..1923
 /organism="unknown"
 /mol_type="genomic DNA"

ORIGIN

Alignment Scores:
 Pred. No.: 3,99e-133 Length: 1923
 Score: 2237.00 Matches: 432
 Percent Similarity: 99.77% Conservative: 1
 Best Local Similarity: 99.54% Mismatches: 1
 Query Match: 99.55% Indels: 0
 DB: 6 Gaps: 0

US-09-985-689A-1-COPY (1-434) x AR368118 (1-1923)

Qy 1 AsnAspValAlaArgGlyIleValLysAlaAspValAlaGlnSerSerTyrGlyLeuTyr 20
 Db 619 AATGATGTTGGCGTGGAAATTTGCAACGGATGTGGCTCAGAGCAGCTACGGTGTGAT 678
 Qy 21 GlyGlnGlyGlnIleValAlaAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40

Db 679 GGCACAGGACAGATCGTAGCGGTGCGGATACAGGCGCTTGATACAGGTGCGCAATGACAGT 738
 QY 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
 Db 739 TCGATCGATGAAGCCTTCGCGGGAATTAATCTGCAATATATGCAATGGAGCGAGAT 798
 QY 61 AsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
 Db 799 AATGCCAATGATACGAATGGTCTATGGTAGCGATGGCTGGCTGGCTCGGTATTAGCAACGCG 858
 QY 81 SerThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100
 Db 859 TCCACTAATAAGGAATGGCGCTCAGGGCAATCTAGTCTTCAATCTATCATGATAGC 918
 QY 101 GlyGlyGlyLeuGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyrSer 120
 Db 919 GGTGGGGGACTTGGAGGACTTACCTCGAATCTGCAACCTTATTGAGCCAGCATACAGT 978
 QY 121 AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAsnGlyValAlaTyrThr 140
 Db 979 CTGTGTGCCGAATTCATCAAACTCTGGGAGCGACGAGTGAATGGGCTTACACAAC 1038
 QY 141 AspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAla 160
 Db 1039 GATTCAGAAATGTGGATGACTATGTGCGCAAAATGATGACGATCCTTTTGGCTGCC 1098
 QY 161 GlyAsnGlyLeuProAsnGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 180
 Db 1099 GGAATGAGAGCGGACCGGCAACCATCAGTGCACAGGACAGCTAAATAATCAATA 1158
 QY 181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 200
 Db 1159 ACAGTCGGAGTACGGAAACCTCCGCCCAAGCTTTGGTCTTATGCGGACAAATCAAC 1218
 QY 201 HisValAlaGlnPheSerArgGlyProThrLysAspGlyArgIleLysProAspVal 220
 Db 1219 CATGTGGCAGATTCCTTACGTGGACCGCAAGAGATGGACGATCAAAACCGGATGTC 1278
 QY 221 MetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 240
 Db 1279 ATGGCACCGGGAAGCTTACATCTATCAGCAAGATCTTCTTCCACCGGATTCCTCCTC 1338
 QY 241 TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyThrSerMetAlaThrProIle 260
 Db 1339 TGGCGCAACCATGACAGTAAATATGATATGATCATGGGTGGAACTCCATGCTACACGATC 1398
 QY 261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPro 280
 Db 1399 GTTGCTGGAAACGTGGCACAGCTTCGTGAGCATTTTGTGAATAACAGAGGCATCACCA 1458
 QY 281 LysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspIleGlyLeuGlyTyr 300
 Db 1459 AAGCCTTCTCTATTAAAGCGGCACCTGATGCGGCTGCGAGCTGCATCGGCCCTTGGTAC 1518
 QY 301 ProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 320
 Db 1519 CCGAAGCTTAACAGAGATGGGACGAGTGACATGGATTAATCTCCAGACCTTCCCTAT 1578
 QY 321 ValAsnGlySerSerSerLeuSerThrSerGlnLysAlaThrTyrSerPheThrAlaThr 340
 Db 1579 GTGAACGAGTCCAGTCTCTATCCACAGCCAAAGACGACGTACTCGTTTACTGCTACT 1638
 QY 341 AlaGlyLysProLeuLysIleSerLeuValTrpSerAspAlaProAlaSerThrThrAla 360
 Db 1639 GCGCGCAAGCCTTTGAATATCCCTGGTATGGTCTGATGCGCCCTGGCAGCAACTGCT 1698
 QY 361 SerValThrLeuValAsnAspLeuAsnLeuValIleThrAlaProAsnGlyThrGlnTyr 380
 Db 1699 TCGGTAAAGCTTCAATGATCTGGACCTTGTCAATTCAGCTCCAAATGGCACACAGTAT 1758
 QY 381 ValGlyAsnAspPheThrSerProTyrAsnAspAsnTrpAspGlyArgAsnAsnValGlu 400
 Db 1759 GTAGGAATGACTTTTACTTCGCGCATACATGATAACTGGGATGGCGGCAATAACGTAGAA 1818

QY 401 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 420
 Db 1819 AATGTAATTAATGATGACCAACAGCGGAGCTATACAAATTGAGTACAGGCTTATAAC 1878
 QY 421 ValProValGlyProGlnThrPheSerLeuAlaIleValAsn 434
 Db 1879 GTACCGTTGGACACAGAACTTCTCGTTGGCAATTGTGAAT 1920
 RESULT 6
 A8046403 1920 bp DNA linear BCT 10-MAY-2002
 LOCUS Bacillus sp. 9860 PROA gene for protease, complete cds.
 DEFINITION A8046403
 VERSION A8046403.2 GI:20521152
 SOURCE
 ORGANISM Bacillus sp. 9860
 Bacillus sp. 9860
 Bacteria, Firmicutes; Bacillales; Bacillaceae; Bacillus.
 REFERENCE 1
 AUTHORS Saeki, K., Okuda, M., Hatada, Y., Kobayashi, T., Ito, S., Takami, H. and Horikoshi, K.
 TITLE Novel oxidatively stable subtilisin-like serine proteases from alkaliphilic Bacillus spp.: enzymatic properties, sequences, and evolutionary relationships
 JOURNAL Biochem. Biophys. Res. Commun. 279 (2), 313-319 (2000)
 MEDLINE 20568675
 PUBMED 11118284
 REFERENCE 2 (bases 1 to 1920)
 AUTHORS Saeki, K.
 DIRECT SUBMISSION
 SUBMITTED (20-JUL-2000) Katsuhisa Saeki, Kao corporation,
 Biological Science Laboratory, Ichikaimachi Akabane 2606, Hage,
 Tochigi 321-3497, Japan (E-mail: 387185@kstanet.kao.co.jp,
 Tel: 81-285-68-7400, Fax: 81-285-68-7403)
 COMMENT On May 9, 2002 this sequence version replaced gi:12381938.
 FEATURES
 source
 1..1920
 /organism="Bacillus sp. 9860"
 /mol_type="genomic DNA"
 /strain="9860"
 /db_xref="taxon:133778"
 gene
 1..1920
 /genes="PROA"
 CDS
 1..1920
 /genes="PROA"
 /codon_start=1
 /transl_table=11
 /product="protease"
 /protein_id="BA521266.2"
 /db_xref="GI:20521153"
 /translation="MRKKVFLSVLSAAAILSTVALNPSAGDARTFDLDFKGIQTTT
 DVSGFKRQTGAALFVSENVKLLKLETPANNKLIHVGNPILLETQK
 LETTGAKILDYIPDAVIYVEGDVQSKVRSIEHVESVPYLPKYKIDPOLPTKGAST
 LVKALADTKQNNKVLQIGIEIAQYASNDVHYITAKPEYKVNNDVARGIVKADVA
 QSSYGLYGGQGIYVAVADTGLDTRGNDSSNHEAFRGKITALYALGRTNANDTNGHTS
 VAGSVNGATNKMAPQANLVFQSDSSGGLGGLPSNLQTLFQSFAGARIHTNS
 WGAAVNGATYDTSRNVDDYVRKNDMTILFAAGNERNGGTISAPGAKNAITVGATEN
 LRPSGYSADININHAQFSRGPTKQGRIPDVMAPGTIYLSARSLAPDSFWANHD
 SKTYNGMGTSMATPIVAGNVAQLREHFVKRGITTPKPSLLKALIAAADVGLYGVNG
 NQGWGVTDKSLNAVYNVESALSTSQTATVTATAGKPLKISLWSDAPASTTAS
 VTLVNDLDLIVTAPNCTRYVGNDFSPFNNNDGRNVNFINSPQSQSYTILEVQAY
 NVFVGPQNFSIAIVN"

ORIGIN

Alignment Scores:
 Pred. No.: 2,22e-129 Length: 1920
 Score: 2178.00 Matches: 417
 Percent Similarity: 99.31% Conservative: 14
 Best Local Similarity: 96.08% Mismatches: 3
 Query Match: 96.93% Indels: 0
 DB: 1 Gaps: 0

US-09-985-689A-1-COPY (1-434) x AB046403 (1-1920)

QY 1 AsnAspValalaArgGlyLeuValAlaAspValAlaGlnSerSerTyrGlyLeuTyr 20
Db 616 AATGATGTGCCAGAGATATTGCAAAAGCGATGTGCCACAGACAGCTACGGTTGTAT 675
QY 21 GlyGlnGlyGlnLeuValAlaAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db 676 GGCACAGCCAGATTGTCCGAGTTGCCGATCTGGATTGGATACAGAAAGACGACGT 735
QY 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
Db 736 TCGATGCATGAAGCTTCCCGGTAAATAACAGCAGCATATATGCATGGGTCCGACGAT 795
QY 61 AsnAlaAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
Db 796 AATGCGAATGATACGAACGCTGATGATCCCATGTGGCAGGTTCGGTATTAGAAATGCG 855
QY 81 SerThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100
Db 856 GCAACGAATAAGGAATGGCACCTCAAGCGAATCTGGTTTTTCAATCCATCATGGATAGC 915
QY 101 GlyGlyGlyLeuGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyrSer 120
Db 916 AGTGGTGGGTGGAGCTTGCCTTCCATCTGCAACCTTATTCAGCCAGCATTCAGT 975
QY 121 AlaGlyAlaArgIleHisThrAsnSerTyrGlyValAlaAlaValAsnGlyAlaTyrThr 140
Db 976 GCAGGTGCCAGATTTCATACAACTCTGGGGGCGCGGTGAATGGGCGCTACAGCACA 1035
QY 141 AspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAla 160
Db 1036 GATTCCAGAAATGGGATGACTATGAAGAAATAATGATATGACGATTTCTTTTCGGGGT 1095
QY 161 GlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 180
Db 1096 GGGATGAAAGCGCGAAGCGCGTACCATCGTCACCTGGTACGCTAAAGCGCAT 1155
QY 181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 200
Db 1156 ACAGTCGGCGCAACCGAAACCTCGCGTCCAAAGCTTCGGTTCTTCATGCAGATAATATAAC 1215
QY 201 HisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVal 220
Db 1216 CAGCTTCACAGTTCTCTTCCCGTGGCCGACAAAGATGGCGAATCAAGCCTGATGTC 1275
QY 221 MetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerPhe 240
Db 1276 ATGGCGCAGGACATACATTTTATCAGCAAGATCTTCTTCGACCCGATTCCTCCTTC 1335
QY 241 TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIle 260
Db 1336 TGGCGGAATCATGACAGCAATATGCCATATATGGTGGACGCTCCATGGCAACCGCAT 1395
QY 261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPro 280
Db 1396 GTTGGCGGGAATGTGCACAGCTCCGAGACATTTGTGAAATAATAGAGGAATCACTCCT 1455
QY 281 LysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspIleGlyLeuGlyTyr 300
Db 1456 AAGCTTTCCCTATTGAAAGCAGCTTTGATTGCAGGTGCTCTCATGTGGATTGGGTTAT 1515
QY 301 ProAsnGlyAsnGlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 320
Db 1516 CCGAACCGGAACCCAGGATGGGCGGAGTGCCTCGATTAATTCGTTGAACGTTCCTAT 1575
QY 321 ValAsnGluSerSerLeuSerThrSerGlnLysAlaThrTyrSerPheThrAlaThr 340
Db 1576 GTGAACCAATCCAGTGCCTTCAACTAGCCAAAAGCGACATATACCTTTTACTGCAAGC 1635
QY 341 AlaGlyLysProLeuLysIleSerLeuValTyrSerAspAlaProAlaSerThrThrAla 360
Db 1636 CGGGCAAGCCATGGAATAATCTCCTGGTATGTTCGATGCCCTCCGACGACACTTGCT 1695

QY 361 SerValThrLeuValAsnAspLeuAsnLeuValIleThrAlaProAsnGlyThrGlnTyr 380
Db 1696 TCTGTAAACCTGGTCAATGATTTGGATTGGTTCATTACAGCACCAACGGAAGATAT 1755
QY 381 ValGlyAsnAspPheThrSerProTyrAsnAspAsnTyrAspGlyArgAsnValGlu 400
Db 1756 GTCCGGATGACTTCTCAGCACCATTTGACAAATPACTGGGATGCCCGCAATACGTAGAA 1815
QY 401 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 420
Db 1816 AATGTATTATTATTAATTCGCCCAAGTGGCAACATATACCATTTAGGTGCAAGCATATAAT 1875
QY 421 ValProValGlyProGlnThrPheSerLeuAlaIleValAsn 434
Db 1876 GTGCGGTTGGACCACAAAACCTTCTCGTTGGCAATTCGTGAAC 1917
RESULT 7
AR368116 1920 bp DNA linear PAT 12-SEP-2003
LOCUS Sequence 3 from patent US 6376227.
DEFINITION AR368116
ACCESSION AR368116
VERSION AR368116.1 GI:34601777
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1920)
AUTHORS Takaiwa, M., Okuda, M., Saeki, K., Kubota, H., Hitomi, J., Kageyama, Y.,
Shikata, S., and Nomura, M.
TITLE Alkaline protease
JOURNAL Patent: US 6376227-A 3 23-APR-2002;
FEATURES Location/Qualifiers
source 1..1920
/organism="unknown"
/mol_type="genomic DNA"
ORIGIN
Alignment Scores:
Pred. No.: 2,22e-129 Length: 1920
Score: 2178.00 Matches: 417
Percent Similarity: 99.31% Conservative: 14
Best Local Similarity: 96.08% Mismatches: 3
Query Match: 96.93% Indels: 0
DB: 6 Gaps: 0
US-09-985-689A-1-COPY (1-434) x AR368116 (1-1920)
QY 1 AsnAspValAlaArgGlyLeuValAlaAspValAlaGlnSerSerTyrGlyLeuTyr 20
Db 616 AATGATGTGCCAGAGATATTGTCAGCGGATCTGGCACAGACGACGTACGGTTGTAT 675
QY 21 GlyGlnGlyGlnLeuValAlaAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db 676 GGCACAGCCAGATTGTCCGAGTTGCCGATCTGGATTGGATACAGAAAGACGACGT 735
QY 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
Db 736 TCGATGCATGAAGCTTCCCGGTAAATAACAGCAGCATATATGCATGGGTCCGACGAT 795
QY 61 AsnAlaAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
Db 796 AATGCGAATGATACGAACGCTCATGCCATGCCATGTGGCAGGTTCGGTATTAGAAATGCG 855
QY 81 SerThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100
Db 856 GCAACGAATAAGGAATGGCACCTCAAGCGAATCTGGTTTTTCAATCCATCATGGATAGC 915
QY 101 GlyGlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyrSer 120
Db 916 AGTGGTGGGTGGAGGCTTGCCTTCCATCTGCAACCTTATTCAGCCAGCATTCAGT 975
QY 121 AlaGlyAlaArgIleHisThrAsnSerTyrGlyValAlaAlaValAsnGlyAlaTyrThr 140

```

Db      976 GCAGGTGCCAGATTTCATACAACTCTGGGGGAGCGGTGAATGGGGCTTACAGACA 1035
Qy      141 ApSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleuPheAlaLa 160
Db      1036 GATTCCAGAAATGTGATGACATATGTAAGGAAATATGATGACGATCTTTTCGGGCT 1095
Qy      161 GlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaLe 180
Db      1096 GGGANTGAAGCGGACCGGTACCATCAGTCACCTGGTACCGCTAAACACGCCATA 1155
Qy      181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 200
Db      1156 ACAGTCGGCGCACACCGAATCTCGCTCAAGCTTCGGTTCCTATGCGATATTAATTAAC 1215
Qy      201 HisValAlaGlnPheSerArgGlyProThrLysAspGlyArgIleLysProAspVal 220
Db      1216 CAGTTTGCAGATTCTCTCCCGTGGCCGACAAAAGATGGCGGAATCAAGCTCATGTC 1275
Qy      221 MetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 240
Db      1276 ATGGCGCCAGGGACATACATATTTATCAGCAAGATCTCTCTTGCAACCGATTCCTCCTTC 1335
Qy      241 TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIle 260
Db      1336 TGGCGCAATCATGACAGCAATATGCTATATGGGTGGACGTCCTCAATGGCAACACCGATT 1395
Qy      261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPro 280
Db      1396 GTTGGCGGGAATGTTGCACAGCTCCGTGAGCATTTTGTGAAATAAGAGGAATCACTCT 1455
Qy      281 LysProSerLeuLeuLysAlaAlaLeuIleAlaGlyValAlaAspIleGlyLeuGlyTyr 300
Db      1456 AAGCTTCTCCATTTGAAGACAGCTTTGATTTGGAGGTGCTGCTGATTTGGATGGGTTAT 1515
Qy      301 ProAsnGlyAsnGlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 320
Db      1516 CCGAACCGAAACCAAGATGGGCGGAGTGACCTGGATAAATCGTGAACGTGGCTAT 1575
Qy      321 ValAsnGlySerSerLeuSerThrSerGlnLysAlaThrTyrSerPheThrAlaThr 340
Db      1576 GTGACCAATCCAGTGGCTTATCACTAGCAAAAGCGACATATACCTTTATCGCAACG 1635
Qy      341 AlaGlyLysProLeuLysIleSerLeuValTrpSerAspAlaProAlaSerThrAla 360
Db      1636 GCGGGCAAGCAATGAAATCTCCCTGGTATGTCGGATGCCCTCGCAAGCACTACTGCT 1695
Qy      361 SerValThrLeuValAsnAspLeuAsnLeuValIleThrAlaProAsnGlyThrGlnTyr 380
Db      1696 TCTGTAAACCTGGTCAATGATTGGATTGGTTGGTCATTACAGCACCAACGAAACAGATAT 1755
Qy      381 ValGlyAsnAspPheThrSerProTyrAsnAspAsnTrpAspGlyArgAsnAsnValGlu 400
Db      1756 GTCGGGAATGACTTCTCAGACCATTTGACAAATACTGGGATGGCGCAATAACGTAGAA 1815
Qy      401 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 420
Db      1816 AATGTATTTATTAATTCGCCCAAGTGGAAACATATATACCATTTGAGTGGCAAGCATATAAT 1875
Qy      421 ValProValGlyProGlnThrPheSerLeuAlaIleValAsn 434
Db      1876 GTGCGGTTGGACCAACAACTTCTCGTTGGCAATTTGTGAC 1917

```

RESULT 8

```

AB046406
LOCUS      AB046406
DEFINITION Bacillus sp. NV1 PROE gene for protease, partial cds.
ACCESSION  AB046406
VERSION    AB046406.1 GI:12381944
KEYWORDS   Bacillus sp. NV1
SOURCE     Bacillus sp. NV1
ORGANISM   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.

```

```

REFERENCE
AUTHORS   1 (sites)
           Saeki, K., Okuda, M., Hatada, Y., Kobayashi, T., Ito, S., Takami, H. and
           Horikoshi, K.
TITLE      Novel oxidatively stable subtilisin-like serine proteases from
           akaiphilic Bacillus spp.: enzymatic properties, sequences, and
           evolutionary relationships
JOURNAL    Biochem. Biophys. Res. Commun. 279 (2), 313-319 (2000)
MEDLINE    20568675
PUBMED     11118284
REFERENCE  2 (bases 1 to 1302)
           Saeki, K.
AUTHORS   Direct Submission
TITLE      Submitted (20-JUL-2000) Katsuhisa Saeki, Kao corporation,
           Biological Science Laboratory, Ichikamachi Akabane 2606, Haga,
           Tochigi 321-3497, Japan (E-mail:1387185@kastanet.kao.co.jp,
           Tel:81-285-68-7400, Fax:81-285-68-7403)
FEATURES   Location/Qualifiers
            source
              1..1302
                /organism="Bacillus sp. NV1"
                /mol_type="genomic DNA"
                /strain="NV1"
                /db_xref="taxon:133781"
            gene
              1..1302
                /gene="PROE"
            CDS
              <1..>1302
                /genes="PROE"
                /codon_start=1
                /transl_table=11
                /product="protease"
                /protein_id="BAB21269.1"
                /db_xref="GI:12381945"
                /translations="NDVARGIVKADVQAQSSYGLYGQGVVAVADTGLDTRNDSSMHE
                AFSGKTAIVAGRTNNANDPNHGTHVAGSVLNGTSNKGMAPOANLVQSVYDSDNG
                GLGGLPNSVTLSQAYSGARITHNSGAPVNGAYTDSRVNDDYVRKNDMAVLFAA
                DNAPGPFILUSASLAPDSSFWANHDSKAYMGTSMATPIVAGNVAQRERFIKNR
                GITPKSLLKALTAGADTGLGYPGNOGWRVTLDRKSLNVAFAVNETSLSTNQKAT
                YSFTAQSGKPLKISLVMSDAPASTASVTLNLDLIVITAPNGTKYGVNDFTPAYDNN
                WDGRNNVENVFINAPQSGTYTVEQVAYNVYVQGPQAFSLAIV"
ORIGIN
Alignment Scores:
Pred. No.:      4,75e-127      Length:      1302
Score:          2138.00      Matches:    405
Percent Similarity: 97.93%      Conservative: 20
Best Local Similarity: 93.32%      Mismatches: 9
Query Match:    95.15%      Indels:    0
DB:             1           Gaps:        0
US-09-985-689A-1-COPY (1-434) x AB046406 (1-1302)
Qy      1 AsnAspValAlaArgGlyIleValLysAlaAspValAlaGlnSerSerTyrGlyLeuTyr 20
Db      1 AACCATGTTGCCAGAGCATTTGTAAGCCGATGTTGCCAGAGCAGCATATGTTTATAT 60
Qy      21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db      61 GGGCAAGGCGCAAGTGGTTCAGTAGCGGATACCGGACTGGATACAGGCGGTATGACAGC 120
Qy      41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
Db      121 TCGATGCATGAAGCGTTCCGAGGAAGATTACCGCGATATATGCCCTTGGGAAGACAAAC 180
Qy      61 AsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
Db      181 AACGCCAATGATCAAAACGGACACGGACCGCATGTTCCCGATCTGTTTAGGAAACCGT 240
Qy      81 SerThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100
Db      241 ACTTCGAATAAGGGATGGCTCCACAGCTAACCTAGTTTCCATCTGTATGACAGC 300
Qy      101 GlyGlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyrSer 120

```

301 AATGGCGAGCTTGGCGAGCTTCCATTAAGTACATATTATTCAGCCAGGCATATAGT 360
121 AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAsnGlyAlaThrThr 140
361 GCCGGTCCAGAAATCCATCAAGAACTCATGGGAGCCCTGTAACAGGAGCGTACATCT 420
141 AspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAla 160
421 GATTCAGAAACGTAGACGATTATGTCGTAATAATGATATGGCGGTCTTTTGGAGCG 480
161 GlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 180
481 GGTAAACGAGGCGGAATGGCGGACAAATCAGTCTCTCTGTCACCGAAGATGCTATC 540
181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 200
541 ACAGTAGGGGACACAGAAAACCTCGCGCCAGGTTTGGATCTTATGCTGACACATCAAT 600
201 HisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVal 220
601 CATGTAGCACAGTTTCTCTCCGCGGACCTACAAAGGATGGAGCTATCAAAACGGACGTA 660
221 MetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 240
661 ATGGCACAGAACTTATTTATCGGCAAGATCTTCTTTGGCTCCGGACTCCTCATTC 720
241 TrpAlaAsnHisAspSerLysTyrAlaThrMetGlyGlyThrSerMetAlaThrProIle 260
721 TGGCAAAACCATGACAGCAAAATGCTTATATGGTGGAACTATCCATGGCGACACCGATT 780
261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPro 280
781 GTAGCTGGTAACTGTCACAGTACGFGAACATTTTCATCAAAACAGAGAAATCACTCCT 840
281 LysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspIleGlyLeuGlyTyr 300
841 AAACCATCTTCTGTAAGCAGCTCTTATTTGCCGAGCACTCATATCGCTCTTGGCTAT 900
301 ProAsnGlyAsnGlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsnValAlaThr 320
901 CCGAGTGGAAACCAAGATGGGAGAGTAACATGGACAAGTCACTTAATGTAGCTTTC 960
321 ValAsnGluSerSerSerLeuSerThrSerGlnLysAlaThrTyrSerPheThrAlaThr 340
961 GTAAATGAACACAGCTTCTTATCTACTAACCAAAAGCTACGTATTCATTTACTGCACAA 1020
341 AlaGlyLysProLeuLysIleSerLeuValTrpSerAspAlaProAlaSerThrThrAla 360
1021 TCAGGCAAACTTTGAAGATTTTCATTCGTTGGTCTGATGCACCGGCAAGTACTTCGCA 1080
361 SerValThrLeuValAsnAspLeuAsnLeuValIleThrAlaProAsnGlyThrGlnTyr 380
1081 TCGGTTACATGGTGAATGATCTGGATCTGGTGATTAACGCTCCAATGGAAAGATAT 1140
381 ValGlyAsnAspPheThrSerProTyrAsnAspAsnTrpAspGlyArgAsnAsnValGlu 400
1141 GTTGGAAACGACTTTCGTCTCCATGATAATAACTGGGATGGACGTAACATGTAGAG 1200
401 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaThrAsn 420
1201 AACGTGTTTTCATATGTCGCAAGGCGGAACGTATACATGAGGTACAGGCTTACAAT 1260
421 ValProValGlyProGlnThrPheSerLeuAlaIleValAsn 434
1261 GTTCCCAAGGCGGAGGCGTCTTCTTTGGCTATTGTGAAC 1302

RESULT 9

AR069954
LOCUS 3003 bp DNA linear PAT 18-FEB-2000
DEFINITION Sequence 41 from patent US 5891701.
ACCESSION AR069954
VERSION AR069954.1 GI:7220842
KEYWORDS

SOURCE Unknown.
ORGANISM Unknown.
REFERENCE Unclassified.
1 (bases 1 to 3003)
AUTHORS Sloma,A. and Christianson,L.
TITLE Nucleic acids encoding a polypeptide having protease activity
JOURNAL Patent: US 5891701-A 41 06-APR-1999;
FEATURES Location/Qualifiers
source 1..3003
/organism="unknown"
/mol_type="unassigned DNA"
Alignment Scores:
Pred. No.: 1.74e-125 Length: 3003
Score: 2120.50 Matches: 405
Percent Similarity: 97.93% Conservative: 20
Best Local Similarity: 93.32% Mismatches: 8
Query Match: 94.37% Indels: 1
DB: 6 Gaps: 1
US-09-985-689A-1-COPY (1-434) x AR069954 (1-3003)
QY 1 AsnAspValAlaAlaArgGlyIleValLysAlaAspValAlaGlnSerSerTyrGlyLeuTyr 20
Db 1470 AATGACGTGGCGCGGTCGCAATTGTGAAGCAGACGTCGCACAAATAACTTTGGCTTATAT 1529
QY 21 GlyGlnGlyGlnIleValAlaAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db 1530 GGACAAGGACAGATTGTAGCAGTTGCTGATATCGGCTTGATACAGGAAGAAATGACAGT 1589
QY 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
Db 1590 TCGATGTCATGAAGCATTCGCGGTAAAGATTACCCGACTATATGACCTGGGAGAACCAAT 1649
QY 61 AsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
Db 1650 AACGCCAATGATCCAAATGGACATGGAAACCCATGTTGCTGGATCTGTGTAGGAAT-- 1706
QY 81 SerThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100
Db 1707 GCTACAAATAAAGGAGTGGCACCCGACCAAGCCCAATCTAGTCTTTCAATCTATTATGGATAGT 1766
QY 101 GlyGlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyrSer 120
Db 1767 GGTGAGGCGCTGGGAGGACTACCTGCTTAATCTACAAATATTTCAGTCAAGCATATAGT 1826
QY 121 AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAsnGlyAlaTyrThrThr 140
Db 1827 GCTGAGCGAGAAATTCATACGAATTCATGGGCGCTCCAGTAAACGGTGCCTATACGACA 1886
QY 141 AspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAla 160
Db 1887 GACTCTCGAAATGTGTATGATATGTAGAGAAAATGATATGACGATCTTTTTCGGGCC 1946
QY 161 GlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 180
Db 1947 GGAAATGAGGACCCAGGTAGCGGTACAAATCAGTCACCGAGACAGCAAAATAATGCGATT 2006
QY 181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 200
Db 2007 ACAGTTGGGCAACCGAAACCTACGTCCAAAGTTCGGATCTTATCGGATTAATATTAAAC 2066
QY 201 HisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVal 220
Db 2067 CATGTTGCTCAATCTCTTCACGAGGCTCTACTAGAGATGGACGTAATTAAGCCGACGTC 2136
QY 221 MetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 240
Db 2127 ATGGCACAGGTAGCTATATTTCTCTGCTAGATCATCATTAGCTCCAGATTCTCTCATTC 2186
QY 241 TrpAlaAsnHisAspSerLysTyrAlaThrMetGlyGlyThrSerMetAlaThrProIle 260

```

Db 2187 TGGGCAACACCATGATAGTAAATATGCTACATGCTGGTGGTACTTCTATGGCTACTCCAATT 2246
Qy 261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValIleAsnArgGlyLeuThrPro 280
Db 2247 GTAGCAGGTAATGTTGCACAATTAAGGGAGCATTTTGTGAAATAATAGAGGGGTAACTCCT 2306
Qy 281 LysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspIleGlyLeuGlyTyr 300
Db 2307 AAGCCTTCCTTTTAAAGCTGCTTTAATTGTCAGTGGTGGGAGTTGGACTTGGCTTT 2366
Qy 301 ProAsnGlyAsnGlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 320
Db 2367 CCAATATGGTAACCAAGGATGGGAGAGTAAGCTTAGATAATCCCTAAATGTGTCATTT 2426
Qy 321 ValAsnGluSerSerSerLeuSerThrSerGlnLysAlaThrTyrSerPheThrAlaThr 340
Db 2427 GTGATGAAACGAGCCCTTTATCAACAGTCAAAAGCAACATATTCGTTACGGCTCAA 2486
Qy 341 AlaGlyLysProLeuLysIleSerLeuValTyrSerAspAlaProAlaSerThrThrAla 360
Db 2487 GCTGGTAAACCCCTTAAATAATATCACTTGTGTCAGATGCCACCAAGTAGCAGCGCA 2546
Qy 361 SerValThrLeuValAsnAspLeuLeuValIleThrAlaProAsnGlyThrGlnTyr 380
Db 2547 TCACCTAACCTTAGTGAATGATTAGACTTAGTAATCACTGCACCAATGGAACTAAATAC 2606
Qy 381 ValGlyAsnAspPheThrSerProTyrAsnAspAsnTyrAspGlyArgAsnAsnValGlu 400
Db 2607 GTGCGAATGACCTTACAGCACCGTATGATAACAATGGGATGGCAGAAACAACGTGGA 2666
Qy 401 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 420
Db 2667 AATGCTTTATCAATGCTCTCTAAGCGGAAGTATACAGTCCGAAGTCGAGCTTACAAT 2726
Qy 421 ValProValGlyProGlnThrPheSerLeuAlaIleValAsn 434
Db 2727 GTACCAGTAAGTCGCAACCTTTCTTAGCGAATGTACAT 2768

RESULT 10
LOCUS BD062155 3003 bp DNA linear FAT 27-AUG-2002
DEFINITION Nucleic acids encoding a polypeptide having protease activity.
ACCESSION BD062155
VERSION BD062155.1 GI:22607760
KEYWORDS JP 2001514529-A/39.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Sloma,A. and Christianson,L.
Nucleic acids encoding a polypeptide having protease activity
Patent: JP 2001514529-A 39 11-SEP-2001;
NOVO NORDISK BIOTECH INC
PN JP 2001514529-A/39
PD 11-SEP-2001
PF 09-JUN-1998 JP 1999503145
PR 12-JUN-1997 US 08/873479
PI ALAN SLOMA,LYNNE CHRISTIANSON
PC C12N15/57,C12N15/75,C12N9/54,C12K14/00
CC Strandedness: Single;
CC Topology: Linear;
FH Key Location/Qualifiers.
FEATURES
source
1..3003
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
ORIGIN
Alignment Scores:
Pred. No.: 1.74e-125 Length: 3003
Score: 2120.50 Matches: 405

```

```

Percent Similarity: 97.93% Conservative: 20
Best Local Similarity: 93.32% Mismatches: 8
Query Match: 94.37% Indels: 1
DB: 6 Gaps: 1
US-09-985-689A-1-COPY (1-434) x BD062155 (1-3003)
Qy 1 AsnAspValAlaArgGlyIleValLysAlaAspValAlaGlnSerSerTyrGlyLeuTyr 20
Db 1470 AATGACGTGGCCCGTGGCATTGTGAAGCAGACGTCGCACAAATAACTTTGGCTTATAT 1529
Qy 21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db 1530 GGAACAAGGACAGATTGTAGCAGTTGCTACTGGCTTGTATACAGAGAAATGACAGT 1589
Qy 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
Db 1590 TCGATGATGAGACATTCGCGGTAGATTACCGCACTATATGCACTGGGCGAGACCAAT 1649
Qy 61 AsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
Db 1650 AACGCCAATGATCCAAATGGACATGGAAACCCATGTTGCTGGATCTGTGTAGGAAT--- 1706
Qy 81 SerThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleValAspSer 100
Db 1707 GCTACAAATTAAGGATGGCACCGCAAGCAATCTAGTCTTTCATCTATTATGGATAGT 1766
Qy 101 GlyGlyGlyLeuGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyrSer 120
Db 1767 GGTGGAGGCGTGGAGGACTACCTGCTAATACAAACATTTATTCAGTCAAGCATATAGT 1826
Qy 121 AlaGlyAlaArgIleHisThrAsnSerTyrGlyAlaAlaValAsnGlyAlaTyrThrThr 140
Db 1827 GCTGGAGCGAGAAATTCATCGAATTCATGGGGGCTCCAGTAAACGGTGCCCTATACGCA 1886
Qy 141 AspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAla 160
Db 1887 GACTCTCGAAATGTTGATGATTATGTGAGAAAAAATGATATGACGATCTTTTGGGGCC 1946
Qy 161 GlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAlaIle 180
Db 1947 GGAATAAGAGGACAGGACGCGTACAATCAGTGCACAGGACAGCAAAAATGCGATT 2006
Qy 181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 200
Db 2007 ACAGTTGGGCAACCCGAAACCTACCTCCAGCTTCGGATCTTATGCGATAATATTAAC 2066
Qy 201 HisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVal 220
Db 2067 CATGTTGCTCAATCTCTTCACGAGGTCCTACTAGAGATGGACGTATTAAAGCGGACGTC 2126
Qy 221 MetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerPhe 240
Db 2127 ATGGCACCGAGTACGTATATTCTCTGCTAGATCATCATTAAGTCCAGATTCCTCATTC 2186
Qy 241 TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyThrSerMetAlaThrProfile 260
Db 2187 TGGGCAACACCATGATAGTAAATATGCTACATGGGTGGTACTTCTATGGCTACTCCAAT 2246
Qy 261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValIleAsnArgGlyIleThrPro 280
Db 2247 GTAGCAGGTAATGTTGCACAATTAAGGGAGCATTTTGTGAAATAATAGAGGGGTAACTCCT 2306
Qy 281 LysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspIleGlyLeuGlyTyr 300
Db 2307 AAGCCTTCCTTTTAAAGCTGCTTTAATTGTCAGTGGTGGGAGTTGGACTTGGCTTT 2366
Qy 301 ProAsnGlyAsnGlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 320
Db 2367 CCAATATGGTAACCAAGGATGGGAGAGTAAGCTTAGATAATCCCTAAATGTGCGATTT 2426
Qy 321 ValAsnGluSerSerSerLeuSerThrSerGlnLysAlaThrTyrSerPheThrAlaThr 340

```

Db 2427 GTGAATGAACGAGCCCTTTTCAACAAAGTCAAAAAGCAACATATTCTGTTTACGGCTCAA 2486

Qy 341 AlaglyysProLeuylleSerLeuValTrpSerAspAlaProAlaSerThrAla 360

Db 2487 GCTGGTAAACCTTAAATAATACATTGTTGTGATGATCCACAGTACGACGCGCA 2546

Qy 361 SerValThrLeuValAsnAspLeuAsnLeuValThrAlaProAsnGlyThrGlnTyr 380

Db 2547 TCACCTAACTTTAGTGAATGATTAGACCTTAGTAACTCACTGCACCAAAATGGAACATAATAC 2606

Qy 381 ValGlyAsnAspPheThrSerProTyrAsnAspAsnTrpAspGlyArgAsnValGlu 400

Db 2607 GTCGAAATGACITTCAGCACCGTATGATACAAATTTGGATGGCAGAACCAACGTTGGA 2666

Qy 401 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 420

Db 2667 AATGTGTTTATCAATGCTCCTCAAGCGGAACGTATACAGTCAAGTGCAGGCTTACAAT 2726

Qy 421 ValProValGlyProGlnThrPheSerLeuAlaIleValAsn 434

Db 2727 GTACCAATAAGTCGCAACCTTTTCTTTAGCGATTGTACAT 2768

RESULT 11

AB046405 1299 bp DNA linear BCT 23-JAN-2001

LOCUS

DEFINITION Bacillus sp. SD521 PROD gene for protease, partial cds.

ACCESSION AB046405

VERSION AB046405.1 GI:12381942

KEYWORDS

SOURCE

ORGANISM

Bacillus sp. SD521

Bacillus sp. SD521

Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.

REFERENCE

1 (sites)

1 (sites) Okuda, M., Hatada, Y., Kobayashi, T., Ito, S., Takami, H. and Horikoshi, K.

AUTHORS

Novel oxidatively stable subtilisin-like serine proteases from alkaliphilic Bacillus spp.: enzymatic properties, sequences, and evolutionary relationships

JOURNAL

Biochem. Biophys. Res. Commun. 279 (2), 313-319 (2000)

MEDLINE

20568675

PubMed

11118284

REFERENCE

2 (bases 1 to 1299)

1299

1299

Direct Submission

AUTHORS

Saeki, K.

TITLE

Submitted (20-JUL-2000) Katsuhisa Saeki, Kao corporation, Biological Science Laboratory, Ichikaimachi Akabane 2606, Haga, Tochigi 321-3497, Japan (E-mail:3818@kstanet.kao.co.jp, Tel:81-285-68-7400, Fax:81-285-68-7403)

JOURNAL

Location/Qualifiers

FEATURES

source

1..1299

/organism="Bacillus sp. SD521"

/mol_type="genomic DNA"

/strain="SD521"

/db_xref="taxon:133780"

1..1299

/gene="PROD"

<1..>1299

/gene="PROD"

/codon_start=1

/transl_table=11

/product="protease"

/protein_id="BAB21268.1"

/db_xref="GI:12381943"

/translation="NDVARGIVKADVQNNYGLYGQGVVAVADTGLDTRNDSSMHE AFRTGKIALVALGRNTNNDPNGHTVAGSVLGNALNKMAPQANLVFQSIIDSSGG LGGLPSNLTLFSGAWNAGAR IHTNSWGPVNGAYTANGSRQDEYVRNNDMTVLPAAG NEGNCSNTISAPGTAKNAITVGATENVRPSGLADNPNHIAOFSSRGATROGRIPKD VTAPGTILSARSLAPDSSFWMYNSKYAMGTSMTAPIVAGNVAQLREHFLKRG ITPKPSLIKALAGATDVGIFGSPQDGQWGRVLDKSLNVAIVNATLALATGCKATY SFQQAQKPLKISLWTDAPGSTASVTLVNDLDTLTPNGQKYVNDPFSYPYDNN DGRNNVENVFINAPQSGTYTIEVQAVNPSPQRFSLAIHV"

gene

CDS

ORIGIN

Alignment Scores:

Pred. No.: 7,1e-118 Length: 1299

Score: 1993.50 Matches: 380

Percent Similarity: 94.24% Conservative: 29

Best Local Similarity: 87.56% Mismatches: 24

Query Match: 88.72% Indels: 1

DB: 1 Gaps: 1

US-09-985-689A-1-COPY (1-434) x AB046405 (1-1299)

Qy 1 AsnAspValAlaAArgGlyIleValIysAlaAspValAlaGlnSerSerTyrGlyLeuTyr 20

Db 1 AATGATGTAGCAGAGGAGTATGTAAGCAGACGTTGCAACAAACATATACGGCTATAT 60

Qy 21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40

Db 61 GGCACAGGTCAAGTAGTTGCAGTAGCGATACGGGTTTAGATACAGGTCGTAAACGATAGT 120

Qy 41 SerMetHisGluAlaPheArgGlyIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60

Db 121 TCTATGCATGAAGCATTCCTGGGAAATTAACAGCTCTTTACGCGTTAGGAGAACTAAC 180

Qy 61 AsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80

Db 181 AATGCAATGATCCGAATGGGCATGGTACGCATGTAGCAGGTTCTGTACTTTGGTAAT--- 237

Qy 81 SerThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100

Db 238 GCTTTAAATAAAGGAATGGCTCCGCAAGCTAACTTAGTCTTTCCAATCTATTATGGATAGC 297

Qy 101 GlyGlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyrSer 120

Db 298 AGCGAGGATTAGGTGGATTAACCATCGAATTTGAATCGTTATTATTAGTCAAGCTTGGAA 357

Qy 121 AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAsnGlyAlaTyrThrThr 140

Db 358 GCTGGGGCTAGAAATTCATACTTTCAGTAAATGGAGCGGTACACTGCT 417

Qy 141 AspSerArgAsnValAlaAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAla 160

Db 418 AACTCGACACAGTGGATGATGTTTCGAAATATGATATGACGGTACTTTTTCGACGA 477

Qy 161 GlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 180

Db 478 GGTAAATGAAGTCTTAATTCAGGAACAATATTAGTCTCCAGGCACAGCAAAAATGCCATT 537

Qy 181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 200

Db 538 ACGGTCCGGCAACAGAAACTATCGCCGAGCTTCGGTTTCATTAGCAGATAACCCAAAT 597

Qy 201 HisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVal 220

Db 598 CATATCGCACAAATTTTCATCAAGAGGAGCTACGAGAGATGACGAAATTAACACGACGTA 657

Qy 221 MetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 240

Db 658 ACAGCTCCTCGGAACATTTATTATTCAGCACGCTTCTTCCTTAGCCCCAGACTCTTCGTT 717

Qy 241 TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIle 260

Db 718 TGGGCGAATTATACAGTAAGTATGCGTACATGGCGGTACTCTATGGCCGACACCTATA 777

Qy 261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPro 280

Db 778 GTTGGGGGAATGTGCGCAATTTACGCGAGCATTTTATAAAAAATAGAGGAATACACCT 837

Qy 281 LysProSerLeuLeuAlaAlaLeuIleAlaGlyAlaAlaAspIleGlyLeuGlyTyr 300

Db 838 AAACTTCTCTTAATTAAGCTGCACCTATCGCTGGGCTACTGATGTGGTTTAGGATAT 897

Qy 301 ProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 320

Db 898 CCAAGTGGTGACCAAGGCTGGGGCGCTGTACTCTAGATAAATCGTTAAATGTAGCGTAT 957

QY 301 ProAsnGlyAsnGlnGlyTTPGlyArgValThrLeuAspLysSerLeuAsnValAlaThr 320
 Db 898 CCAAGTGGTGAACAGAGCTGGGGCGGTGTTACITTTAGATAAATCGITTAATGTAGCGTAT 957
 QY 321 ValAsnGluSerSerLeuSerThrThrSerGlnLysAlaThrTyrSerPheThrAlaThr 340
 Db 958 GTCAATGAAGCAACTGCATTAACAACAGGACAAAGCAACGATTTCGTTCCAAACGCCAA 1017
 QY 341 AlaGlyLysProLeuLysLysSerLeuValTTPSerAspAlaProAlaSerThrThrAla 360
 Db 1018 GCGGGTAACACATTAATAAATCTCGTTAGTAGGACAGATGCACCTCGAAGTACACAGCA 1077
 QY 361 SerValThrLeuValAsnAspLeuAsnLeuValThrAlaProAsnGlyThrGlnTyr 380
 Db 1078 TCTTATACACTAGTTAATGATTAGATCTAGTTACTAGTCTCGAATGACACAAAATAT 1137
 QY 381 ValGlyAsnAspPheThrSerProTyrAsnAspAsnTTPAspGlyArgAsnValGlu 400
 Db 1138 GTAGGTAATGATTAGTTATCTTATGATGATAAATTTGGATGCTCGCAACATGTTGAG 1197
 QY 401 AsnValPheLeuAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaThrAsn 420
 Db 1198 AACGTATTTATAACGCTCGCAATCTGGAACGTATACATTCAGGTTCAAGCGTATAC 1257
 QY 421 ValProValGlyProGlnThrPheSerLeuAlaIleValAsn 434
 Db 1258 GTTCCATCTGACCACAGCGTTTCTCACTAGCTATCGTACAT 1299

RESULT 13
 AB046404
 LOCUS AB046404 1299 bp DNA linear BCT 23-JAN-2001
 DEFINITION Bacillus sp. Y PROC gene for protease, partial cds.
 ACCESSION AB046404
 VERSION AB046404.1 GI:12381940
 KEYWORDS
 SOURCE Bacillus sp. Y
 ORGANISM Bacillus sp. Y
 Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 REFERENCE 1 (sites)
 AUTHORS Saeki,K., Okuda,M., Hatada,Y., Kobayashi,T., Ito,S., Takami,H. and Horikoshi,K.
 TITLE Novel oxidatively stable subtilisin-like serine proteases from alkaliphilic Bacillus spp.: enzymatic properties, sequences, and evolutionary relationships
 JOURNAL Biochem. Biophys. Res. Commun. 279 (2), 313-319 (2000)
 MEDLINE 20568675
 PUBMED 11118284
 REFERENCE 2 (bases 1 to 1299)
 AUTHORS Saeki,K.
 TITLE Direct Submission
 JOURNAL Submitted (20-JUL-2000) Katsuhisa Saeki, Kao corporation, Biological Science Laboratory, Ichikaimachi Akabane 2606, Haga, Tochigi 321-3497, Japan E-mail:387185@kasanet.kao.co.jp, Tel:81-285-68-7400, Fax:81-285-68-7403)
 FEATURES
 Location/Qualifiers
 1..1299
 /organism="Bacillus sp. Y"
 /mol_type="genomic DNA"
 /strain="Y"
 /db_xref="taxon:133779"
 1..1299
 /gene="PROC"
 <1..1299
 /gene="PROC"
 /codon_start=1
 /evidence=experimental
 /transl_table=11
 /product="protease"
 /protein_id="BAB21267.1"
 /db_xref="GI:12381941"
 /translation="NDVARGIVKADVAQNNGYLGQGVVAVADTGLDTRNDSSMHE AFRKIKTALVALGRTNNSDPNGHGVHAGSVLGNALNKGMAPOANLVQSDMDSSGG LGLPLSNLNTLFSQANNAGARIJHTNSKGAIPVNGAYTANSRQVDEYVRNDMTVLVFAAG

NEGNSGTISAPGTAKVAITVGATENRRPFSGIADNPNIHIAQFSSRGATRDGRIKPD
 VTASGTILSARSLAPDSFMYNYSKYVMGTSMTATPVAQNVLAQLRHFHFKNG
 ITPSPILKALAGATDVLGCPNGQGRVTLNKSINVAYNEATATAGSKATY
 SFOAQGRPKISLWTFADPGSTTASTILVNDLVLVTAPNGQKVGNDFSYPIDNWW
 DGRNNVENFINAPQSGTYIIEVOAYNVPSQPFSIAIVH"

Alignment Scores:
 Pred. No.: 3.55e-117 Length: 1299
 Score: 1982.50 Matches: 378
 Percent Similarity: 94.01% Conservative: 30
 Best Local Similarity: 87.10% Mismatches: 25
 Query Match: 88.23% Indels: 1
 DB: 1 Gaps: 1

US-09-985-689a-1-COPY (1-434) x AB046404 (1-1299)

QY 1 AsnAspValAlaArgGlyIleValLysAlaAspValAlaGlnSerSerTyrGlyLeuTyr 20
 Db 1 AATGATGTAGCAGAGGATAGTAAAGCTGATGTTGCACAAAACAATACGGATTATAT 60
 QY 21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
 Db 61 GGACAGGTCAAGTAGITGCGTAGCGGACACAGGCTAGATACAGTCTGTAACGATAGT 120
 QY 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
 Db 121 TCTATGATGAAGCATTCGCGGGAAATACACGCTCTTTACGCGTTAGGAAGAACTAAT 180
 QY 61 AsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
 Db 181 AATGGAGTATCCGATGGCATGGCCACATGATAGCAGGTTCTGTACTTGGTAAT --- 237
 QY 81 SerThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100
 Db 238 GCTTTAAATAAAGGAATGGCTCCGCAAGCTAACTAGTCTTCCAACTATTATTATGATAGC 297
 QY 101 GlyGlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyrSer 120
 Db 298 ACCGAGGATTTAGTGGCTTACCATCGAATTAATAGTATTATTAGTCAAGCTTGAAT 357
 QY 121 AlaGlyAlaArgIleHisThrAsnSerTTPGlyAlaAlaValAsnGlyAlaTyrThrThr 140
 Db 358 GCTGGAGCAGAAATTCATCTCTTGGGAGCCCGAGTAAATGGAGCGTACACTGCT 417
 QY 141 AspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAla 160
 Db 418 AACTCGACAGCAAGTGGATGAATATGTCGAATATATGATATGACGGTACTTTTTCAGCT 477
 QY 161 GlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 180
 Db 478 GGTAATGAAGTCTCTTAATTCAGGAACAATTAGTCTCCAGGTACAGCGNAAAATGCTATT 537
 QY 181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 200
 Db 538 ACGGTGGCGCAACGAAACATATCGCCCAAGCTTCGCTCGATAGCAGATAACCCAAAT 597
 QY 201 HisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVal 220
 Db 598 CATATTGCACATTTTCATCCAGAGGAGTACAGAGGATGGACCAATTAAGCCTGACGTA 657
 QY 221 MetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 240
 Db 658 ACAGCTCTCTGGAACATTTATTTATCAGCACGTTCTTCTTCTTAGCTCCAGACTCTTCGTT 717
 QY 241 TTPAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIle 260
 Db 718 TGGCGCAATTATACAGTAATATGCGGTATATGGCGGTACTCTCATGGCGACACCTATT 777
 QY 261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPro 280
 Db 778 GTTGACGGGAATGTGCGCAATTACGTGAGCATTTTATAAAAAATAGAGGTATTACTCT 837


```

QY 281 LysProSerLeuLeuLysAlaAlaLeuLysAlaAlaAspGlyLeuGlyTyr 300
DB 838 ALGCTTCTTTTAAATAAACTGCATTCCTGCTGCTACTGATGTTGGTTAGGATAT 897
QY 301 ProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 320
DB 898 CCTAATGGTGACCAAGCTGGGGGGTGTACTCTAAATAAATCGTTAAATGTAGCGTAT 957
QY 321 ValAsnGluSerSerSerLeuSerThrSerGlnLysAlaAlaThrTyrSerPheThrAlaThr 340
DB 958 GTCAATGAAGCACTTCATAGCCACAGGACAAAAGCAACGTATTCGTTCACGACAA 1017
QY 341 AlaGlyLysProLeuLysLysSerLeuValTrpSerAspAlaProAlaSerThrThrAla 360
DB 1018 GCGGGTAAACCTTTAAATCTCGTTAGTATGACAGATGCTCTCGGAATACAACTGCA 1077
QY 361 SerValThrLeuValAsnAspLeuAsnLeuValThrAlaProAsnGlyThrGlnTyr 380
DB 1078 TCTTATACACTAGTTAATGATTTAGATCTAGTATTACTGCTCCGAATGGACAAAATAT 1137
QY 381 ValGlyAsnAspPheThrSerProTyrAsnAspAsnTrpAspGlyArgAsnValGlu 400
DB 1138 GTAGGAATGATTTTAGTTATCCTTATGATAATACTGGATGTCGCAACAATGTTGAG 1197
QY 401 AsnValPheLeuAsnAlaProGlnSerGlyThrTyrThrThrLeuValGlnAlaTyrAsn 420
DB 1198 AACGTATTTATAAACCTCGCAATCTGGAACGTATATAATTTGAGTTCAAGCGTATAAT 1257
QY 421 ValProValGlyProGlnThrPheSerLeuAlaAlaLeuValAsn 434
DB 1258 GTACCATCTGGCCACAGCGTTCTCACTAGCTATCGTACAT 1299

RESULT 14
E03808
LOCUS      E03808
DEFINITION DNA encoding alkaline protease ya.
ACCESSION E03808
VERSION    E03808.1  GI:2172022
KEYWORDS   JP 1992197182-A/1.
SOURCE     Bacillus sp.
ORGANISM   Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
REFERENCE  1 (bases 1 to 2218)
AUTHORS    Tobe, S., Odera, M. and Asai, Y.
TITLE      DNA CODING ALKALINE PROTEASE YA ENZYME AND PRODUCTION OF ALKALINE
           PROTEASE YA USING THE DNA
JOURNAL    Patent: JP 1992197182-A 1 16-JUL-1992;
           LION CORP
COMMENT    OS Bacillus sp.
           PN JP 1992197182-A/1
           PD 16-JUL-1992
           PF 28-NOV-1990 JP 1990327110
           PI TOBE S, ODERA M, ODERA MOTOYASU, ASAI YOSHIO
           PC C12N15/57, C12D3/386, C12N9/54, (C12N15/57, C12R1.07), (C12N9/54,
           PC C12R1.07);
           CC strandedness: Double;
           CC topology: Linear;
           CC hypothetical: No;
           CC anti-sense: No;
           CC *source: strain=y strain;
           CC *source: clone=pUB8A;
           FH Key Location/Qualifiers
           FH CDS 218..2218
           FT /product='precursor of alkaline protease ya'
           FT sig_peptide 218..823
           FT mat_peptide 824..2212
           FT /product='Precursor of alkaline protease ya'.
           FT Location/Qualifiers
           FT 1..2218
           FT /organism='Bacillus sp.'
           FT /mol_type='genomic DNA'
           FT /db_xref='taxon:1409'
FEATURES
source

```

```

ORIGIN
Alignment Scores: 7.99e-117 Length: 2218
Pred. No.: 1981.50 Matches: 378
Score: 94.01% Conservative: 30
Percent Similarity: 94.01%
Best Local Similarity: 87.10% Mismatches: 25
Query Match: 88.18% Indels: 1
DB: 6 Gaps: 1

US-09-985-689a-1-COPY (1-434) x E03808 (1-2218)
QY 1 AsnAspValAlaAlaArgGlyLeuValLysAlaAspValAlaGlnSerSerTyrGlyLeuTyr 20
DB 824 AATGATCTAGCAAGAGGATAGTAAAGCTGATGTTGCAAAACAACTACCGATTATAT 883
QY 21 GlyGlnGlyGlnLeuValAlaAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
DB 884 GGACAGGTCUACTAGTTGCGATGGGACACAGGCTTAGATACAGGTGTAACGATAGT 943
QY 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
DB 944 TCTATGATCATGAAGCAATTCGCGGAAAAATCACAGCTCTTTACGCTTAGGAGACTAAT 1003
QY 61 AsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
DB 1004 AATGCGAGTGATCCGAATGGGATGGCACAATGATGAGGTTCTGTACTTGGTAAAT --- 1060
QY 81 SerThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100
DB 1061 GCTTTAAATAAGGAATGGCTCCGCAAGCTAACTTAGTCTTCCAATCTATTATGGATAGC 1120
QY 101 GlyGlyLeuGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyrSer 120
DB 1121 AGCGGAGGATTAGTGCTTACCATCGAATTAATACGTTATTATTAGTCAAGTCTGGAAT 1180
QY 121 AlaGlyAlaArgIleHisThrAsnSerTrpGlyValAlaValAsnGlyAlaTyrThr 140
DB 1181 GCTGAGCAAGATTCATCTACTCTTGGGAGCCCAAGTAATGGAGCGTACACTGCT 1240
QY 141 AspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAla 160
DB 1241 AACTCGAGACAAGTGGATGATGTTGCGAAATTAATGATATGCGTACTCTTTTTCAGCT 1300
QY 161 GlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 180
DB 1301 GGTAAATGAAGTCTTAATTCAGGACAAATTAGTGTCTCCAGGTACAGGAAATGCTATT 1360
QY 181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 200
DB 1361 ACGTCTGGCGCACACGAAAACTATCGCCCAAGCTTCGGTTTCGATAGCAGATAACCCAAAT 1420
QY 201 HisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVal 220
DB 1421 CATATTCACAAATTTTCATCGAGAGGAGCTACGAGGATGGACGAATTAAGCCTGACGTA 1480
QY 221 MetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 240
DB 1481 ACAGCTCTCGAACATTTATTTTATCAGCACAGTCTTCTTCTAGCTCCAGACTCTCGTTT 1540
QY 241 TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIle 260
DB 1541 TGGCGGAATTATACAGATAAATACGCGTATATGGCGGTACCTCCATGGCGACACCTATT 1600
QY 261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPro 280
DB 1601 GTTCAGGGAATGTCGCGCAATTACGTGAGCAATTTTATAAAAAATAGAGGTATTACTCCT 1660
QY 281 LysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspIleGlyLeuGlyTyr 300
DB 1661 AAGCTTTCTTTAAATAAAGCTGCATCTATCGCTGGTCTACTGATGTTGTTTAGGATAT 1720
QY 301 ProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 320

```

```

Db 1721 CCTAGTGGTACCACGCTGGGGCGTGTACTCTAGATAAATCGTTAAATGTAGCGTAT 1780
Oy 321 ValAsnGluSerSerSerLeuSerThrSerGlnIysAlaThrTyrSerPheThrAlaThr 340
Db 1781 GTCAATGAACCACTGCATTAGGCACAGGACAAAGAACGATTCGTTCCAGACACAA 1840
Oy 341 AlaGlyIysProLeuLysIleSerLeuValTyrSerAspAlaProAlaSerThrThrAla 360
Db 1841 GCGGTAACCTTTAAAAATCTCGTTAGTAGGACAGATCTCTCTGGAAGTACAACTGCA 1900
Oy 361 SerValThrLeuValAsnAspLeuAsnLeuValIleThrAlaProAsnGlyThrGlnTyr 380
Db 1901 TCTATACACTAGTAAATGATTAGATCTAGTATTACTCTCTCCGAATGGACAAATAAT 1960
Oy 381 ValGlyAsnAspPheThrSerProTyrAsnAspAsnTyrAspGlyArgAsnAsnValGlu 400
Db 1961 GTAGGAATGNTTTAGTATCTCTATGATAATAAATCGGATGGTGGCAACATGTTGAG 2020
Oy 401 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 420
Db 2021 AACGTATTTAAACGCTCCGCAATCTGGAACGTATATAATTGAGGTTCAAGCGTATAAT 2080
Oy 421 ValProValGlyProGlnThrPheSerLeuAlaIleValAsn 434
Db 2081 GTACCACTGCCCAACGCGTTTCTCACTAGCTATCGTACAT 2122

RESULT 15
AF268611/c 60006 bp DNA linear BCT 30-MAY-2001
LOCUS uncultured marine group II euryarchaeote EBAC37F11, BAC clone
DEFINITION 37F11, partial sequence.
ACCESSION AF268611
VERSION AF268611.1 GI:9664575
SOURCE uncultured marine group II euryarchaeote 37F11
ORGANISM Archaea; Euryarchaeota; Marine Group II; environmental samples.
REFERENCE 1 (bases 1 to 60006)
AUTHORS Beja, O., Suzuki, M.T., Koonin, E.V., Aravind, L., Hadd, A.,
Nguyen, L.P., Villacorta, R., Amjadi, M., Garrigues, C.,
Jovanovich, S.B., Feldman, R.A. and Delong, E.F.
TITLE Construction and analysis of bacterial artificial chromosome
libraries from a marine microbial assemblage
JOURNAL Environ. Microbiol. 2 (5), 516-529 (2000)
MEDLINE 21128453
PUBMED 11233160
REFERENCE 2 (bases 1 to 60006)
AUTHORS Beja, O., Suzuki, M.T., Koonin, E.V., Aravind, L., Hadd, A.,
Nguyen, L.P., Villacorta, R., Amjadi, M., Garrigues, C.,
Jovanovich, S.B., Feldman, R.A. and Delong, E.F.
TITLE Direct Submission
JOURNAL Submitted (17-MAY-2000) R & D, Monterey Bay Aquarium Research
Institute, P.O. Box 628, Moss Landing, CA 95039-0628, USA
FEATURES
source
1..60006
/organism="uncultured marine group II euryarchaeote 37F11"
/mol_type="genomic DNA"
/db_xref="taxon:133822"
/clones="BAC 37F11"
/notes="environmental sample"
1484..3178
/notes="37F11#1"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="S-layer protein"
/protein_id="AAF97179.1"
/db_xref="GI:9664576"
/translation="MEPGQTELDIFVQSPSSAIAQGVNIKRVSDGSCQGLEEREI
PVRVSPNFEINASQNFVSEGGYGLAWMENTGMTSHVTILENIPGVNISMQ
SASLSPHEARIPPEIIPPTNDKQPLEIIVHPQMGVSEITVIHNSIPASTP
VLSSGAGSSQAFSLHLSINPSTVTVPDATNQDEYRLILEEGINFETISITNDIAIQ

```

CDS

CDS

CDS

rRNA

rRNA

CDS

CDS

```

VNIGSILPQVSACQFISDVQFQELGRAPLTETVWSCVNGDLDEQTLRTLTFTIASTNTGE
TIPDLGFTTQNRSEFVNISVDQWPEPQQLIRVVGNGGIEVVSIEIAQVSQK
SNWVGASFSATGDLIAITRTNVEYGVNVCILTITSSSTSSYRIERIIVDVEGSOFA
PIVKQNSISDRSLSATIDCNSPFDVDDPNSNTATAIYVESQSILVTNNLMGT
AITTIFGAIYVQNRREAGVQNSIRSKPNTNNKDTQETKETEYIEDMSLIEIE
EMEIEEPVSUVEITVSETDLSPSGRLDITIRQELDPDVEIVDTTISIEERNSKFFD"
3223..3891
/notes="37F11#2"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="unknown"
/protein_id="AAF97180.1"
/db_xref="GI:9664577"
/translation="MQNSNSTSAFYTLDCVSENLDFTFOALLEIRGNHQLWEKLV
NREAFSLTKQGLDEIDDRNKTITWDATLFTCVRLNRSAPIMLGNKRRFGR
PHGTGSLQVILENLSIPQSEGLYSDLIEHLEYLTERCSEMTKHEFLEGKAGL
MLGVITFEYKALRSLMLGSGVMVSRPEELDGGVREAVRHLNALLMAAERAGGLIH
RQHE"
complement(3892..4797)
/notes="37F11#3"
/codon_start=1
/evidence=not experimental
/product="SAICAr synthase"
/protein_id="AAF97181.1"
/db_xref="GI:9664578"
/translation="MGENLYQGVKQVQWSTDDPILFRFTNQISVFDQIIPSLIPRK
GESNRTTAHFWKLVEQGVGCTHLVELNAPDCLVRKVEVFKPGPAIDRAEWVFP
LEFIVRVLGSAWRFORQGLTSLGVSQDAEYGVKLSFFVEVTTKFAFDNRNID
REALQISNTIDQFDSIQVQALKVDEIIEEAAKNGLIHVDGKKEFALGPKRVLV
DTFGTLDLRWDAAEYVNGECIELSKFVTRHINSGHQAEKLARDSGAIDPPIPA
LPESVIEETASLYSMYERLTSQQF"
complement(4846..5508)
/notes="37F11#4"
/codon_start=1
/evidence=not experimental
/product="translation initiation factor 6"
/protein_id="AAF97182.1"
/db_xref="GI:9664579"
/translation="MPDITLDIHGSDQIGIHLASGVNVLFPRLPTTIEKLRQVVG
LELAPISIGGNLIGALLAGNTKGMADVATESDIALTAIGDVVWEGVNTAGNL
LVNVEGCVASPSIPKDLGLDILSEVLGVDIITITGGQDVVGSVGNVSGVLLHPDV
APEVLLIEILGVPVPMVGTVSGPVYVAGICASNNAGLAGSETTGPENLRIDALG
LI"
5597..5745
/product="5S ribosomal RNA"
5816..5811
/product="23S ribosomal RNA"
complement(8888..10192)
/notes="37F11#7; contains Zn-ribbon and TPR-like repeats"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="unknown"
/protein_id="AAF97183.1"
/db_xref="GI:9664580"
/translation="MNECAVCGIILYLAGSPACGSQLRAQDQSPSDAALPTEVPG
LDAAQAWYDLEGTEAPVEQBEAAASLPFGGEGSOTNIRLPGIGSHPEGIPFEL
SNQDSQEESEVQGVHEVIESTNLNPVALETQCHDSPVVASVEIETPLVAPVAD
LPLIEEPVQVFPVPGPIRLTAIPFVEPTVDSLPLENWSIESFGSQDIFTAEEDVIE
VVFSDLETVVHRHGENYQPMKEDSVYSKSLQPLDHPAQLMVDVGTDSARAL
LESQFTSIGQDWRSAARSPQIVARFPQDSGMANNYGIALQLQVAVNQHSNDIEVS
NSAQFEAAIILREAVTNSQSPESYNLSQALLSGREKALLDLMAGNESQRES
NFTNLRAAILAQGLRYDEAKLLVPLQADKLASGNLLKLPAI"
complement(10248..11969)
/notes="37F11#8"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="permease"
/protein_id="AAF97184.1"

```



```
Qy 258 ThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGly 277
Db 16658 ACAGCAGTAGCAGGTGGTTTCAGCATCCCTCGCAGCTGAATATCTGGTGAAGTTGCTGGA 16599
Qy 278 IleThr---ProLysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspIle 296
Db 16598 ATCAACAACCATCTGCAAGCCCTATTAGGCAACTCTCATCAACGGGGCAGAGATTTA 16539
Qy 297 GlyLeu---GlyTyrProAsnGlyAsnGlnGlyTyrGlyArgValThrLeuAspLysSer 315
Db 16538 GGCACACGGACATTCCTAACGCAACAGGAGATGGGGCAAAATGATCTCGAAATAGC 16479
Qy 316 LeuAsnValAla-----TyrValAsnGluSerSerLeu 327
Db 16478 CTGAACTCTCGTCGTCGTCGGGTTCTCTCGAGTCTTTTCAGGATGATGAGCGTGAATT 16419
Qy 328 SerThrSerGlnLysAlaThrTyrSerPheThrAlaThrAlaGlyLysProLeuLysIle 347
Db 16418 CAGCAGGATTCTCATTTGATTTATTCCTTTGACCTTGATGATCAAAGGTTATTGACATT 16359
Qy 348 SerLeuValTTPSerAspAlaProAlaSerThrThrAlaSerValThr-----Leu 364
Db 16358 ACCTTGGCATGGAGCGATGCAGAGCCAGTCGAAATGCTGCACATCTGAATCAGCTCTA 16299
Qy 365 ValAsnAspLeuAsnLeuValIleThrAlaProAsnGlyThrGlnTyrValGlyAsnAsp 384
Db 16298 TTGAACAATTTAGATTTGATTCTTAATTGCACCCGATGGGTCTTCATACTTGGGTAAATGAC 16239
Qy 385 PheThrSerProTyrAsn-----AspAsnTrpAspGlyArgAsnAsnValGluAsn 401
Db 16238 TTCTCTTCTGGGATTTCACAAACAGCGGTTCCGCTGATTAATCTGAACAATATCGAGCGC 16179
Qy 402 ValPheIleAsnAla-----ProGlnSerGlyThrTyrThrIleGluVal----- 416
Db 16178 ATTGCAATCCCTGCGAGGCGGACGACACAGCGTGATTTGGATGGTTACAGTTGAACAT 16119
Qy 417 -----GlnAlaTyrAsnValProVal 423
Db 16118 CGAGGAGGAGTTCTCAGCGCTACAGTATCGTTATT 16083
```

Search completed: March 15, 2004, 21:48:31
Job time : 3658 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: March 15, 2004, 19:16:41 ; Search time 395 Seconds
(without alignments)
4667.640 Million cell updates/sec

Title: US-09-985-689A-1-COPY

Perfect score: 2247
Sequence: 1 NDVARGIVKADVAQSSYGLY.....EVQAYNPVGPQTFSIAIVN 434

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-Q=/cgn2.1/USPTO.spool/SWOP985689/runat_10032004.112806.19350/app_query.fasta_1.583
-DB=N_geneseq_29Jan04 -QFW=faetap -SUFFIX=rng -MINMATCH=0.1 -LOOPEXT=0
-LOOPEXT=0 -UNITS=bts -START=1 -END=1 -MATRIX=biosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFWT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N_Geneseq_29Jan04:*

| | |
|-----|------------------|
| 1: | Geneseqn1980s.* |
| 2: | Geneseqn1990s.* |
| 3: | Geneseqn2000s.* |
| 4: | Geneseqn2001as.* |
| 5: | Geneseqn2001bs.* |
| 6: | Geneseqn2002s.* |
| 7: | Geneseqn2003as.* |
| 8: | Geneseqn2003bs.* |
| 9: | Geneseqn2003cs.* |
| 10: | Geneseqn2004s.* |

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB ID | Description |
|------------|--------|-------------|--------|------------|--------------------|
| 1 | 2242 | 99.8 | 1923 | 2 AAX37278 | Aax37278 Bacillus |
| 2 | 2237 | 99.6 | 1923 | 2 AAX37279 | Aax37279 Bacillus |
| 3 | 2178 | 96.9 | 1920 | 2 AAX37277 | Aax37277 Bacillus |
| 4 | 2120.5 | 94.4 | 3003 | 2 AAV82382 | Aav82382 Bacillus |
| 5 | 1981.5 | 88.2 | 1299 | 2 AAQ27515 | Aaq27515 Alkali-pr |
| 6 | 447.5 | 19.5 | 1977 | 2 AAT85667 | Aat85667 Thermococ |
| 7 | 447.5 | 19.9 | 1977 | 2 AAX05926 | Aax05926 WO9856926 |
| 8 | 411.5 | 18.3 | 1236 | 2 AAX05920 | Aax05920 Hyperther |

| | | | | | |
|----|-------|------|--------|------------|---------------------|
| 9 | 411.5 | 18.3 | 1566 | 2 AAT85668 | Aat85668 Pyrococcus |
| 10 | 411.5 | 18.3 | 1962 | 2 AAT85695 | Aat85695 Pyrococcus |
| 11 | 411.5 | 18.3 | 1962 | 2 AAX05929 | Aax05929 Hyperther |
| 12 | 398.5 | 17.7 | 1977 | 2 AAT85669 | Aat85669 Protease |
| 13 | 372 | 16.6 | 2121 | 4 ABL54900 | Ab154900 T. yonsei |
| 14 | 348 | 15.5 | 2539 | 2 AAT61454 | Aat61454 Streptomy |
| 15 | 348 | 15.5 | 2809 | 2 AAT61455 | Aat61455 DhpA-mel |
| 16 | 336 | 15.0 | 135638 | 7 ABX34289 | ABX34289 S. atrool |
| 17 | 332.5 | 14.8 | 1329 | 6 ABK74643 | Abk74643 Bacillus |
| 18 | 311.5 | 13.9 | 2532 | 2 AAQ29134 | Aaq29134 Encodes R |
| 19 | 307 | 13.7 | 2835 | 2 AAT08141 | Aat08141 Hyperther |
| 20 | 307 | 13.7 | 4765 | 2 AAT08132 | Aat08132 Protease |
| 21 | 307 | 13.7 | 4765 | 2 AAT85670 | Aat85670 Pyrococcus |
| 22 | 307 | 13.7 | 4765 | 2 AAX05921 | Aax05921 WO9856926 |
| 23 | 306.5 | 13.6 | 1560 | 7 ABZ37569 | Abz37569 Streptomy |
| 24 | 306.5 | 13.6 | 59816 | 7 ABZ37516 | Abz37516 Streptomy |
| 25 | 306.5 | 13.6 | 59816 | 7 ABZ37515 | Abz37515 Streptomy |
| 26 | 302.5 | 13.5 | 1859 | 2 AAT85677 | Aat85677 Thermococ |
| 27 | 296.5 | 13.2 | 3413 | 2 AAV72330 | Aav72330 F. balust |
| 28 | 296 | 13.2 | 898 | 2 AAT08131 | Aat08131 Hyperther |
| 29 | 290.5 | 12.9 | 564 | 2 AAT08134 | Aat08134 DNA seque |
| 30 | 290.5 | 12.9 | 564 | 2 AAT85676 | Aat85676 Thermococ |
| 31 | 282 | 12.6 | 3788 | 9 ADD24905 | Add24905 DNA encod |
| 32 | 278 | 12.4 | 3743 | 9 ADD24901 | Add24901 DNA encod |
| 33 | 276 | 12.3 | 1306 | 6 ABL55784 | Ab155784 Bacillus |
| 34 | 276 | 12.3 | 1330 | 6 ABL55787 | Ab155787 Bacillus |
| 35 | 275 | 12.2 | 546 | 4 ABL53453 | Ab153453 T. yonsei |
| 36 | 270.5 | 12.0 | 2273 | 2 AAQ04339 | Aaq04339 Aquaricin |
| 37 | 270.5 | 12.0 | 2274 | 2 AAQ12838 | Aaq12838 Aquaricin |
| 38 | 270.5 | 12.0 | 2274 | 2 AAQ75859 | Aaq75859 Aqualysin |
| 39 | 261 | 11.6 | 4716 | 9 ABQ80437 | Abq80437 CspA codi |
| 40 | 261 | 11.6 | 4740 | 7 ABZ58957 | Abz58957 Group B S |
| 41 | 256.5 | 11.4 | 1152 | 8 ACC85473 | Acc85473 B gibsoni |
| 42 | 256.5 | 11.4 | 1194 | 8 ADA32120 | Ada32120 DNA encod |
| 43 | 256.5 | 11.4 | 10216 | 2 AAT39279 | Aat39279 Transposo |
| 44 | 256 | 11.4 | 1485 | 6 ABK74647 | Abk74647 Bacillus |
| 45 | 254 | 11.3 | 4650 | 6 ABN71526 | Abn71526 Streptoco |

ALIGNMENTS

RESULT 1

AAX37278

ID AAX37278 standard; DNA; 1923 BP.

XX

AC AAX37278;

XX

DT 20-MAR-2003 (revised)

DT 21-JUL-1999 (first entry)

XX

DE Bacillus alkaline protease encoding DNA.

XX

KW Alkaline protease; Bacillus; casein digestion; oleic acid; enzyme;

KW washing composition; oxidising agent; ss.

XX

OS Bacillus sp.

XX

PN WO9918218-Al.

XX

PD 15-APR-1999.

XX

PF 07-OCT-1998; 98WO-JP004528.

XX

PR 07-OCT-1997; 97JP-00274570.

XX

PA (KAOS) KAO CORP.

XX

PI Takaiwa M, Okuda M, Saeki K, Kubota H, Hitomi J, Kageyama Y;

PI Shikata S, Nomura M;

XX

DR WPI: 1999-287736/27.

DR P-PSDB; AAX17088, AAX17090.

XX

Shikata S, Nomura M;
 WPI: 1999-287736/27.
 P-PSDB; AAY17091.
 Alkali protease from *Bacillus* used in washing powders.
 Disclosure; Page 63-68; 71pp; Japanese.

The invention relates to alkaline proteases produced by strains of *Bacillus*. The proteases ability to digest casein is not inhibited by oleic acid and they have a high stability to oxidising agents. The alkaline protease of the invention has the following properties: (a) it is active over the pH range 4-13 and has at least 80% of its optimum activity over the pH range 6-12; (b) after 30 minutes at 40 deg. C it is stable over the pH range 6-11; (c) its isoelectric point is 8.9-9.1; (d) its ability to digest casein is not inhibited by oleic acid; (e) it has a molecular weight about 43,000 by SDS-PAGE. The alkaline proteases can be used as enzymes in washing compositions for use in automatic dishwashers and for washing clothes. The stability to oxidising agents allows the enzyme to be an effective component of washing compositions including bleaches. The present sequence represents an alkaline protease encoding DNA. (Updated on 20-MAR-2003 to correct DR field.)

Sequence 1923 BP; 578 A; 417 C; 474 G; 454 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 1,24e-162 Length: 1923
 Score: 2237.00 Matches: 432
 Percent Similarity: 99.77% Conservative: 1
 Best Local Similarity: 99.54% Mismatches: 1
 Query Match: 99.55% Indels: 0
 DB: 2 Gaps: 0

US-09-985-689A-1-COPY (1-434) x AAX37279 (1-1923)

QY 1 AsnAspValAlaAaGgGlyIleValIysAlaAspValAlaGlnSerSerTyrGlyLeuTyr 20
 DB 619 AATGATGTGGCGGTGAATGTCAAGCGGATGTGGCTCAGACGACCTACGGGTGTAT 678
 QY 21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
 DB 679 GGACAAAGACAGATCGTAGCGGTGGCCGATACAGGGCTTGATACAGGTGCGCAATGACAGT 738
 QY 41 SerMetHisGluAlaPheAtrGlyIleThrAlaLeuTyrAlaLeuGlyValGThrAsn 60
 DB 739 TCGATGCTAGAGCTTCGCGGGAATTAATCTCATATATATGATGGACGCGAGAT 798
 QY 61 AsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
 DB 799 AATGCCAATGATACGAATGGTCATGGTACGCATGTGGCTCGCTCCGTATTAGGAAACGGC 858
 QY 81 SerThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100
 DB 859 TCCACTATAAAGGAATGGCCCTCAGCGCAATCTAGTCTCCAACTATCATCGGATAGC 918
 QY 101 GlyGlyGlyLeuGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyrSer 120
 DB 919 GGTGGGGACCTGGAGGACTACCTTCGAATCTGCAACCTTATTTCAGCAAGCATACAGT 978
 QY 121 AlaglyAlaAaGlyIleHisThrAsnSerTrpGlyAlaAlaValAsnGlyAlaTyrThr 140
 DB 979 GCTGTGCGCAATTCATACAAATCTCCGGGAGCAGCGAATGGGCTTCACAAACA 1038
 QY 141 AspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAla 160
 DB 1039 GATCCAGAAATGTGGATGACTATGTGGCAAAATGATATGACGATCCITTCGTGCC 1098
 QY 161 GlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 180
 DB 1099 CGGAATGAAGACCGAAGCGGGAACCATCATGTGCACGGCAGCAGCTAAATAATGCAATA 1158
 QY 181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 200

DB 1159 ACAGTCGGAGCTACGGAAACCTCCGCCAAAGCTTGGGTCTTATCGGCAATATCAAC 1218
 QY 201 HisValAlaGlnPheSerArgGlyProThrLysAspGlyArgIleLysProAspVal 220
 DB 1219 CATGTGGCACAGTCTCTTCACGTGGACCGAAGGATGACGGATCAACCGGATGTC 1278
 QY 221 MetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 240
 DB 1279 ATGGCACCGGAACGTTTCATCTATCAGCAAGATCTTCTCTTGCACCGGATTCCTCTTC 1338
 QY 241 TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyThrSerMetAlaThrProIle 260
 DB 1339 TGGGGCAACCATGACAGTAATAATGCAATATGTTGGTGGAAAGCTCCATGGCTACCCGATC 1398
 QY 261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPro 280
 DB 1399 GTTGTGGAAACGTGGCACAGCTTCGTGAGCATTTTGTGAAAAACAGAGGCATCACACCA 1458
 QY 281 LysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspIleGlyLeuGlyTyr 300
 DB 1459 AAGCCTTCTATTAAAAAGCGCACTGATTCGGTGCAGCTGACATCGGCTTGGCTAC 1518
 QY 301 ProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 320
 DB 1519 CGRAACGTACCAAGGATGGGACGAGTGACATTTGGATTAATCCTCGACGTTGCTTAT 1578
 QY 321 ValAsnGluSerSerSerLeuSerThrSerGlnLysAlaThrTyrSerPheThrAlaThr 340
 DB 1579 GTGAACGAGTCCAGTCTCTATCCACCAAGCAAGCGACGTACTCGTTTACTGCTACT 1638
 QY 341 AlaGlyLysProLeuLeuLysIleSerLeuValTrpSerAspAlaProAlaSerThrThrAla 360
 DB 1639 GCCGCAAGCCTTTGAAATCTCCCTGGTATGGTCTGATGCCCTCGCAGCACAACTGCT 1698
 QY 361 SerValThrLeuValAsnAspLeuAsnLeuValIleThrAlaProAsnGlyThrGlnTyr 380
 DB 1699 TCCGTAAACGCTTGTCAATGATCTGGACCTTGTTCATTACCGCTCCAAATGGCACAGTAT 1758
 QY 381 ValGlyAsnAspPheThrSerProTyrAsnAspAsnTrpAspGlyArgAsnValGlu 400
 DB 1759 GTAGGAATGACTTTACTTCGCCATPACATGATTAATCTGGGATGGCCGCAATAACGTAGA 1818
 QY 401 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 420
 DB 1819 AATGTATTATTATATGACCACCAAGCGGACGTATACATTAAGTACAGCTTATAAC 1878
 QY 421 ValProValGlyProGlnThrPheSerLeuAlaIleValAsn 434
 DB 1879 GTACCGGTTGGACCAAGAACTTCTCGTTGGCAATTGTGAAT 1920

RESULT 3
 AAX37277
 ID AAX37277 standard; DNA; 1920 BP.
 XX AAX37277;
 AC AAX37277;
 XX 20-MAR-2003 (revised)
 DT 21-JUL-1999 (first entry)
 XX
 DE Bacillus alkaline protease encoding DNA.
 XX
 KW Alkaline protease; Bacillus; casein digestion; oleic acid; enzyme;
 KW washing composition; oxidising agent; ss.
 XX
 OS Bacillus sp.
 XX
 PN WO9918218-A1.
 XX
 XX 15-APR-1999.
 PD
 PF 07-OCT-1998; 98WO-JP004528.
 XX

PR 07-OCT-1997; 97JP-00274570.
 XX (KAOS) KAO CORP.
 XX Takaiwa M, Okuda M, Saeki K, Kubota H, Hitomi J, Kageyama Y;
 PI Shikata S, Nomura M;
 XX WPI; 1999-287736/27.
 DR P-PSDB; AAY17087, AAY17089.
 XX
 PT Alkali protease from *Bacillus* used in washing powders.
 XX
 PS Disclosure; Page 53-58; 71pp; Japanese.
 XX
 CC The invention relates to alkaline proteases produced by strains of
 CC *Bacillus*. The proteases ability to digest casein is not inhibited by
 CC oleic acid and they have a high stability to oxidising agents. The
 CC alkaline protease of the invention has the following properties: (a) it
 CC is active over the pH range 4-13 and has at least 80% of its optimum
 CC activity over the range pH 6-12; (b) after 30 minutes at 40 deg. C it is
 CC stable over the pH range 6-11; (c) its isoelectric point is 8.9-9.1; (d)
 CC its ability to digest casein is not inhibited by oleic acid; (e) it has
 CC molecular weight about 43,000 by SDS-PAGE. The alkaline proteases can be
 CC used as enzymes in washing compositions for use in automatic dishwashers
 CC and for washing clothes. The stability to oxidising agents allows the
 CC enzyme to be an effective component of washing compositions including
 CC bleaches. The present sequence represents an alkaline protease encoding
 CC DNA. (Updated on 20-MAR-2003 to correct DR field.)
 XX
 SQ Sequence 1920 BP; 576 A; 411 C; 473 G; 460 T; 0 U; 0 Other;

Alignment Scores:
 Pred. No.: 4, 25e-158 Length: 1920
 Score: 2178.00 Matches: 417
 Percent Similarity: 99.31% Conservatives: 14
 Best Local Similarity: 96.08% Mismatches: 3
 Query Match: 96.93% Indels: 0
 DB: 2 Gaps: 0

US-09-985-689a-1-COPY (1-434) x AAX37277 (1-1920)

Qy 1 AsnAspValAlaArgGlyIleValLysAlaAspValAlaGlnSerSerTyrGlyLeuTyr 20
 Db 616 AATGATGTGCCAGGATTTGTCAAGCGGATGTGGCAGACAGCAGCTACGGTTTGAT 675
 Qy 21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
 Db 676 GCACAGCGCAGATTTGCCAGTTGCCGATCTGGATTGGATACAGAGAACAGCAGT 735
 Qy 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
 Db 736 TCGATGCATGAAGCTTCCGCGGTAAATAAACAGCACTATATGCACTGGGTGGACGAAT 795
 Qy 61 AsnAlaAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
 Db 796 AATCGAATGATACGAACGTCATGGTACCATTGGCGAGTTCCGATTATAGGAATGGC 855
 Qy 81 SerThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100
 Db 856 GCAACGAATAAGGAATGGCACCTCAAGCGAACTCGTTTTCATCCATCATGGATAGC 915
 Qy 101 GlyGlyClyLeuGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyrSer 120
 Db 916 AGTGGTGGGCTTGAGGCTTGCCTTCCAACTCTGCAAACTTATTCAGCCAAAGCATTCAT 975
 Qy 121 AlaGlyAlaArgIleHisThrAsnSerTrpGlyValAlaValAsnGlyAlaTyrThrThr 140
 Db 976 GCAGTCCAGATTCATACAACTCTCTGGGGGCGAGCGGTGAATGGCGCTACACGACA 1035
 Qy 141 AspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAla 160
 Db 1036 GATTCCAGAAATGTGGATGACTATGTAAAGGAAATGATATGACGATTTCTTTTCGGCGCT 1095

Qy 161 GlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 180
 Db 1096 GGGATGAAGCGCGAAGCGGGTACCATCAGTCGACCTGTAGCGCTAAAGCGCCATA 1155
 Qy 181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 200
 Db 1156 ACAGTCGGCGCAACCGGAAACCTTCGCTCAAGCTTCGTTCTTATGCAGATAATATTAAAC 1215
 Qy 201 HisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVal 220
 Db 1216 CACGTTGCACAGTTCTCTTCCCGTGGCCGACAAAGATGGCGAATCAAGCCGTGATGTC 1275
 Qy 221 MetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 240
 Db 1276 ATGGCGCGAGGACATACATTTTATCAGCAAGATCTTCTCTGCACCCGATTCCTCCTTC 1335
 Qy 241 TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIle 260
 Db 1336 TGGCGGATCATGACAGCAATATGCTATATGGTGGACGTCCTCATGGCAACACCGATT 1395
 Qy 261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPro 280
 Db 1396 GTTCGGGGGAATGTGCACAGCTCCGTGAGCATTTTGTGAAAAATAGAGCAATCACTCCT 1455
 Qy 281 LysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspIleGlyLeuGlyTyr 300
 Db 1456 AAGCTTCCCTATTGAAAGCAGCTTTGATTCGAGTGCTGCTGATGTGTGATGGGTAT 1515
 Qy 301 ProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 320
 Db 1516 CCGAACGGAACCAAGGATGGGCGGAGTGACCCCTGGATAAATCGTTGAAACGTTGCCCTAT 1575
 Qy 321 ValAsnGluSerSerSerLeuSerThrSerGlnLysAlaThrTyrSerPheThrAlaThr 340
 Db 1576 GTGAACGAATCCAGTGCCCTATCAACTAGCCAAAGACGACATATACCTTTACTGCAACG 1635
 Qy 341 AlaGlyLysProLeuLysIleSerLeuValTrpSerAspAlaProAlaSerThrAla 360
 Db 1636 CGGGCAAGCCATTGAAATCTCCCTGCTGATGGTGGATGCCCTGCAAGCACTACTGCT 1695
 Qy 361 SerValThrLeuValAsnAspLeuAsnLeuValIleThrAlaProAsnGlyThrGlnTyr 380
 Db 1696 TCTGTAAACCCGTGCAATGATTTGGATTTGGTCATTACAGCACCAACGGAACAGATAT 1755
 Qy 381 ValGlyAsnAspPheThrSerProTyrAsnAspAsnTrpAspGlyArgAsnAsnValGlu 400
 Db 1756 GTGGGATGACTTCTCAGCACCATTTGACATAAATGCGATGGCGCCCAATAACGTAGAA 1815
 Qy 401 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 420
 Db 1816 AATGTATTATTATTAATTCGCCCAAGTGAACCATATACCATTTGAGGTGCAAGCATATAAT 1875
 Qy 421 ValProValGlyProGlnThrPheSerLeuAlaIleValAsn 434
 Db 1876 GTGGCGGTGGACCAACAAACTTCTCGTTGGCAATTGTGAAC 1917

RESULT 4
 AAV82382
 ID AAV82382 standard; DNA; 3003 BP.
 XX
 AC AAV82382;
 XX
 DT 12-APR-1999 (first entry)
 XX
 DE *Bacillus* JP170 protease gene.
 XX
 KW Protease; detergent; surfactant; leather processing; debittering;
 XX flavour; ss.
 XX
 OS *Bacillus* sp.
 XX
 FH Key Location/Qualifiers
 FT CDS 846..2771

FT /*tag= a
 FT sig_peptide 846..944
 FT /*tag= b
 FT mat_peptide 1470..2768
 FT /*tag= c

FN WO9856927-A2.

PD 17-DEC-1998.

PF 09-JUN-1998; 98WO-US012005.

PR 12-JUN-1997; 97US-00873479.

PA (NOVO) NOVO NORDISK BIOTECH INC.

PI Sloma A, Christianson L;

XX WPI; 1999-080908/07.

DR P-PSDB; AAW89547.

XX Novel protease from *Bacillus subtilis* LC20 - useful in laundry and
 FT dishwashing detergents and for leather processing.

XX Claim 11; Page 52-53; 77pp; English.

CC This nucleotide sequence encodes a novel protease (see AAW82382) of
 CC *Bacillus* sp. JPI170 (NCIB 12513). The sequence in plasmid p170BAN is
 CC contained in *Bacillus subtilis* LC20 NRRL B-21680. The protease gene was
 CC isolated from chromosomal DNA of JPI170 following preparation of probes
 CC based on protease N-terminal and internal peptides (see AAW89549-50),
 CC screening of chromosomal libraries, isolation of the 3' end of the gene
 CC by inverse PCR (see AAW82410-11), reconstruction of 5' and 3' ends and
 CC PCR amplification (see AAW82412-16). Claimed recombinant host cells can
 CC be used in a method for producing the protease. The protease is used in
 CC laundry and dishwashing detergents, for institutional and industrial
 CC cleaning, and for leather processing, as well as for debittering and
 CC enhancing the degree of hydrolysis of protein hydrolysates, for flavour
 CC development through hydrolysis of proteins, degradation of undesired
 CC peptides and in enzymatic synthesis of peptides. It has enhanced
 CC stability towards oxidation under alkaline conditions, e.g. towards
 CC bleaching agents of the peroxy type. The invention also provides mutant
 CC cells in which the protease activity is diminished. Such cells can be
 CC used for the production of heterologous recombinant proteins

XX SQ Sequence 3003 BP; 982 A; 504 C; 645 G; 872 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 1.9e-153 Length: 3003
 Score: 2120.50 Matches: 405
 Percent Similarity: 97.93% Conservative: 20
 Best Local Similarity: 93.32% Mismatches: 8
 Query Match: 94.37% Indels: 1
 DB: 2 Gaps: 1

US-09-985-689A-1-COPY (1-434) x AAW82382 (1-3003)

Qy 1 AsnAspValAlaArgGlyIleValLysAlaAspValAlaGlnSerSerTyrglyLeuTy 20
 Db 1470 AATGACGTGGCCGGCGGATTGTGAAGACGACGTGCGACAAATAACTTTGGCTTATAT 1529
 Qy 21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
 Db 1530 GGACAAGGACAGATTGTAGCAGTTGCTGATCTACCTGGGCTTGATACAGGAAGAAATGACAGT 1589
 Qy 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyAlaLeuGlyArgThrAsn 60
 Db 1590 TCGATGTCATGAAGCATTCGCGGGTAAGATTACCGCACTATATGCTGGGCGAGACCGAAT 1649
 Qy 61 AsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
 Db 1650 AACGCCAATGATCCAAATGGGACATGGAAACCCATGTTGCTGATCTGTGTAGGAAT--- 1706

RESULT 5

Qy 81 SerThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100
 Db 1707 GCTACCAATAAAGGGATGGCACCGCAACCACTAGTCTTTCAATCTATTATGGTAGT 1766
 Qy 101 GlyGlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTySer 120
 Db 1767 GGTGGAGGGCTGGGAGGACTACTCTGTAATCTACAAACATTATTGCTGAGTCAAGCATATAGT 1826
 Qy 121 AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAsnGlyAlaTyThrThr 140
 Db 1827 GCTGGAGCGAGATTCTATACGATTCTATGGGGGCTCCAGTAAACGGTGCCTATACGACA 1886
 Qy 141 AspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAla 160
 Db 1887 GACTCTCGAAATGTTGATGATTATGTGAGAAATAATGATATGACGATCTTTTTCGGGCC 1946
 Qy 161 GlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 180
 Db 1947 GGAAATGAGGACAGGTAGCGGTACATCACTGTCACCGAGAACAGCAAAAAATGCGATT 2006
 Qy 181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyAlaAspAsnIleAsn 200
 Db 2007 ACAATTGGGGCAACCGAAACCTAGCTCCAGCTTCGGATCTTATGCGGATTAATATTAAAC 2066
 Qy 201 HisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleIleProAspVal 220
 Db 2067 CATGTTGCTCAATCTCTTCACGAGTCTCTAGAGATGGACGTATTAAAGCGGACGTC 2126
 Qy 221 MetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 240
 Db 2127 ATGGCACCAAGGTACGTATATTTCTCTGTAGATCATCATAGTCCAGATCTCTCATTC 2186
 Qy 241 TrpAlaAsnHisAspSerLysTyAlaTyMetGlyGlyThrSerMetAlaThrProIle 260
 Db 2187 TGGGCAAAACCATGATAGTAAATATGCTACATGGGTGGTACTTCTATGCTACTCCAAAT 2246
 Qy 261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPro 280
 Db 2247 GTAGCAGGTAATGTTGCACAAATTAAGGAGCATTGTTGAAAAATAGAGGGTAACCTCT 2306
 Qy 281 LysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspIleGlyLeuGlyTy 300
 Db 2307 AAGCCTTCCTCTTTAAAGCTCCTTAATGTCAGTGTCTGGGATGTTGGACTTGGCTTT 2366
 Qy 301 ProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaTy 320
 Db 2367 CCBAATGGTAACCAAGGATGGGAGAGTAACGTAGATAAATCCCTAAATGTCGCAAT 2426
 Qy 321 ValAsnGluSerSerSerLeuSerThrSerGlnLysAlaThrTySerPheThrAlaThr 340
 Db 2427 GTGAATGAACAGAGCCCTTTATCAACAGCTCAAAAAGCAACATATTCGTTTACGGCTCAA 2486
 Qy 341 AlaGlyLysProLeuLysIleSerLeuValTTPSerAspAlaProAlaSerThrThrAla 360
 Db 2487 GCTGTAAACCCCTTAAANAATACCTTGTGTCAGATGCACCAAGTAGCAGCAGCGCA 2546
 Qy 361 SerValThrLeuValAsnAspLeuAsnLeuValIleThrAlaProAsnGlyThrGlnTy 380
 Db 2547 TCCTAACTTAGTGAATGATTAGATTAGTAATCACTGACCAAAATGGAATAAATAC 2606
 Qy 381 ValGlyAsnAspPheThrSerProTyAsnAspAsnTrpAspGlyArgAsnValGlu 400
 Db 2607 GTCGGAATAGACTTTACAGCACCGTATGATAAATTTGGATGGCAGAAACCAACGTGGA 2666
 Qy 401 AsnValPheIleAsnAlaProGlnSerGlyThrTyThrIleGluValGlnAlaTyAsn 420
 Db 2667 AATGTGTTTATCAATGCTCTCAAAGCGGAACGTATACAGTCGAAGTGCAGGCTTACAAT 2726
 Qy 421 ValProValGlyProGlnThrPheSerLeuAlaIleValAsn 434
 Db 2727 GTACCAAGTAAGTCCGCAACCTTTTCTTAGCGATTGACAT 2768

AAQ27516
 ID AAQ27516 standard; DNA; 1299 BP.
 XX
 AC AAQ27516;
 XX
 DT 05-FEB-1993 (first entry)
 XX
 DE Alkali-protease Ya enzyme gene.
 XX
 KW Alkali resistance; surface active agent resistance; detergency improver;
 KW ss.
 OS Bacillus sp. Y.
 XX
 XX
 FH Key Location/Qualifiers
 FT CDS 1..1299
 FT /*tag= a
 XX
 XX JP04197182-A.
 XX
 XX 16-JUL-1992.
 XX
 XX 28-NOV-1990; 90JP-00327110.
 XX
 XX 28-NOV-1990; 90JP-00327110.
 XX
 XX (LLOY) LION CORP.
 XX
 DR WPI; 1992-288440/35.
 DR P-PSDB; AAR26274.
 XX
 PT DNA coding alkali-protease Ya enzyme - has good alkali and surfactant
 PT resistance and improves detergency.
 XX
 PS Claim 3; Page 2; 17pp; Japanese.
 XX
 CC The sequence is that of the alkali-protease Ya enzyme gene which can be
 CC used in the recombinant production of Ya enzyme. Ya enzyme is excellent
 CC in alkali resistance and surface active agent resistance and improves
 CC detergency
 XX
 SQ Sequence 1299 BP; 401 A; 246 C; 297 G; 355 T; 0 U; 0 Other;
 Alignment Scores:
 Pred. No.: 3,41e-143 Length: 1299
 Score: 1981.50 Matches: 378
 Percent Similarity: 94.01% Conservative: 30
 Best Local Similarity: 87.10% Mismatches: 25
 Query Match: 88.18% Indels: 1
 DS: 2 Gaps: 1
 US-09-985-689A-1-COPY (1-434) x AAQ27516 (1-1299)
 Qy 1 AsnAspValAlaArgGlyIleValLysAlaAspValAlaGlnSerSerTyrGlyLeuTyr 20
 Db 1 AATGATAGCAGAGGGATAGTAAAGCTGATGTTGCACAAAACAATTACGGATTATAT 60
 Qy 21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyAAspAspSer 40
 Db 61 GGACAAGGTCACACTAGTTGCAGTAGCGGACACAGGCTTAGATACAGGTCGTACGATAGT 120
 Qy 41 SerMetHisGluAlaPheArgGlyIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
 Db 121 TCTATGATGAAGCATTTCCGCGGAAAATAACACAGCTCTTTACGCGTTAGGAAGACTAAT 180
 Qy 61 AsnAlaAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
 Db 181 AATCGAGTGTCCGANTGGCATGGCACATGATACAGGTTCTGTACTTGGTAAT--- 237
 Qy 81 SerThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100
 Db 238 GCTTTAAATAAAGGAATGGCTCCGCAAGCTAACTAGTCTTCCAACTATTATGGATAGC 297

RESULT 6

AAT85667

ID AAT85667 standard; DNA; 1977 BP.

XX

AC AAT85667;

Qy 101 GlyGlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyrSer 120
 Db 298 ACCGAGGATTAGTGGCTTACCACTCAACTTAATACGTTATTATTAGTCAAGCTTGAAT 357
 Qy 121 AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAsnGlyAlaTyrThr 140
 Db 358 GCTGGAGCAGAAATTCATACCTCTTTGGGAGCCCCAGTAAATGGAGCGTACACTGCT 417
 Qy 141 AspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAla 160
 Db 418 AACTCGAGACAAGTGGATGAGTATGTTCCGAATATGATATGACGGTACTTTTTGCGAGCT 477
 Qy 161 GlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 180
 Db 478 GGTAATGAAGTCTCTTAATTCAGGAACAATTAGTCTCCAGGTACAGCGAAAAATGCTATT 537
 Qy 181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 200
 Db 538 ACGGTCGGCGCAACGGAACCTATCGCCCAAGCTTCGGTTCCGATAGCATTAACCCAAAT 597
 Qy 201 HisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVal 220
 Db 598 CATATTGCACAATTTTCATCGAGAGGAGCTACGAGGATGAGCAATTAAGCCTGACGTA 657
 Qy 221 MetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 240
 Db 658 ACAGCTCCTGGACAATTTATTTATCAGCACGTTCTTCCTTAGCTCCAGACTCTTCGTT 717
 Qy 241 TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIle 260
 Db 718 TGGCGGAATTTAACAAGTAAATACGCTATATGCGCGGTACCTCCATCGGCGACACTATT 777
 Qy 261 VallalaglyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPro 280
 Db 778 GTTCAGGGAATTCGCGCAATTTACGTGAGCAATTTATAAAAAATAGAGTATTACTCCT 837
 Qy 281 LysProSerLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspIleGlyLeuGlyTyr 300
 Db 838 AAGCTTCTTTAATAAAGCTGCACTTATCGCTGCTACTGATGATGTTGGTTAGGATAT 897
 Qy 301 ProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 320
 Db 898 CCTAGTGTGCACCAAGGCTGGGGCGTGTACTCTAGATAAATCGTTAAATGTAGCGTAT 957
 Qy 321 VallanGluSerSerSerLeuSerThrSerGlnLysAlaThrTyrSerPheThrAlaThr 340
 Db 958 GTCAATGAAGCACTGCAATTAGCCACAGGACAAAAGCAACGATTTCTGTTCAAGCAAA 1017
 Qy 341 AlaGlyLysProLeuLysIleSerLeuValTrpSerAspAlaProAlaSerThrAla 360
 Db 1018 GCGGGTAAACCTTTAAAAATCTCGTTAGTAGGACAGATGCTCTCGGAAGTACAACTGCA 1077
 Qy 361 SerValThrLeuValAsnAspLeuAsnLeuValIleThrAlaProAsnGlyThrGlnTyr 380
 Db 1078 TCTTATACACTAGTTAATGATTAGTTAGTATTACTGCTCCGATGGCAAAAAATAT 1137
 Qy 381 ValGlyAsnAspPheThrSerProTyrAsnAspAsnTrpAspGlyArgAsnAsnValGlu 400
 Db 1138 GTAGGAATGATTTTAGTTATCTTATGATAAATAACTGGGATGTCGCAACAATGTTGAG 1197
 Qy 401 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 420
 Db 1198 AACGATTATTAACCGCTCCGCAATCTGGAACGATATATAATTGAGTTCGAGCGGTATAAT 1257
 Qy 421 ValProValGlyProGlnThrPheSerLeuAlaIleValAsn 434
 Db 1258 GTACCATCTGCCCCACAGCGTTTCTCACTAGTATCGTACAT 1299

XX 17-OCT-2003 (revised)
 DT 20-APR-1998 (first entry)
 XX Thermococcus protease coding sequence.
 XX Protease; research reagent; thermal stability; thermococcus celer; ss.
 XX Thermococcus celer; DSM-2476.
 XX WO9721823-A1.
 XX 19-JUN-1997.
 XX 07-NOV-1996; 96WO-JP003253.
 XX 12-DEC-1995; 95JP-00323285.
 XX (TAKI) TAKARA SHUZO CO LTD.
 XX Takakura H, Morishita M, Yamamoto K, Mitta M, Asada K;
 XX Tsunashawa S, Kato I;
 DR WPI; 1997-332794/30.
 DR P-PSDB; AAW24121.
 XX Protease(s) and genes encoding them obtained from Thermococcus and
 PT Pyrococcus strains - have extremely high thermal stability and are useful
 PT industrially and as research reagents.
 XX Claim 3; Page 86-87; 159pp; Japanese.
 CC This sequence represents the coding sequence for the protease from
 CC Thermococcus celer DSM-2476. This sequence encodes a protease of the
 CC invention. The proteases of the invention have extremely high thermal
 CC stability. The proteases can be used as research reagents, and
 CC industrially in the food, drug and chemical industries. (Updated on 17-
 CC OCT-2003 to standardise OS field)
 XX SQ Sequence 1977 BP; 453 A; 659 C; 554 G; 311 T; 0 U; 0 Other;

Alignment Scores:
 Pred No.: 4.76e-25 Length: 1977
 Score: 447.50 Matches: 137
 Percent Similarity: 44.66% Conservative: 68
 Best Local Similarity: 29.85% Mismatches: 153
 Query Match: 19.92% Indels: 101
 DB: 2 Gaps: 18

US-09-985-689A-1-COPY (1-434) x AAT85667 (1-1977)

Qy 8 ValLysAlaAspValAlaGlnSerSerTyrrGlyLeuTyrrGlyGlnGlyGlnIleValAla 27
 Db 433 ATAGGGCCGATACCGTCTGGAACCTCCCTCGGCTACGACGAAAGCGTGTGGTGGC 492
 Qy 28 ValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisGluAlaPheArg 47
 Db 493 ATCTGATACGGGTATAGACGGAC-----CACCCCGATCTGGAAG 534
 Qy 48 GlyLysIleThrAlaLeuTyrr---AlaLeuGlyArgThrAsnAsnAlaAsnAspThrAsn 66
 Db 535 GGCAAGGTCATAGCTGGTACGACCGCGCTCAACGGCAGGTCGACCCCTACGATGACCAG 594
 Qy 67 GlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGlySerThrAsnLys----- 84
 Db 595 GGACACGGAAACCCACAGCTGCGGGTATCGTTCGCCGGAACCGGCGATTAACCTCCAGTAC 654
 Qy 85 ---GlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMet-----AspSerGly 101
 Db 655 ATAGCGCTGCCCGCGCGAAGCTCGTGGGTCAAGGTTCTCGGTGCGGACGGTTCG 714
 Qy 102 GlyClyLeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyrrSerAla 121

RESULT 7

AAK05926

ID AAK05926 standard; DNA; 1977 BP.

Db 715 GGAAGCGTCTCCACCATCATCGGGGTGTTGACTGGGTCTCCAGAACAGGACAAGTAC 774
 Qy 122 GlyAlaArgIle-----HisThrAsnSer 129
 Db 775 GGGATAAGGTCATCAACCTCTCCCTCGGCTCTCCAGAGCTCCGACGGAACCGACTCC 834
 Qy 130 TrpGlyAlaAlaValAsnGlyAlaTyrrThrThrAspSerArgAsnValAspAspTyrrVal 149
 Db 835 CTCAGTCAGGCGCTCAACACCGCTGGGACGCC----- 867
 Qy 150 ArgLysAsnAspMetThrIleLeuPheAlaAlaGlyAsnGluGlyProAsnGlyGlyThr 169
 Db 868 -----GGTATAGTAGTCTCGGTGCGCGCAACAGCGGCGCAACACCTACACC 918
 Qy 170 IleSerAlaProGlyThrAlaLysAsnAlaIleThrValGlyAlaThrGluAsnLeuArg 189
 Db 919 GTCCGCTCACCGCGCGCGAGCAAGGTCAACCGTCGGTCA----- 963
 Qy 190 ProSerPheGlySerTyrrAlaAspAsnIleAsnHisValAlaGlnPheSerSerArgGly 209
 Db 964 -----GTTGACAGCAACCAACATCGCCAGCTTCTCCAGCAGGGA 1005
 Qy 210 ProThrLysAspGlyArgIleLysProAspValMetAlaProGlyThrPheIleLeuSer 229
 Db 1006 CCGACCGCGGACGGAAGCTCAAGCGGAAGTCTGCGCCCGCGGTGACATCATAGCC 1065
 Qy 230 AlaArgSerSerLeuAlaProAspSerSerPheTrpAlaAsnHisAspSerLysTyrrAla 249
 Db 1066 CCGCGCGCCAGC-----CGAGCTGGACCCCGCAAGCAAGCTACTACACC 1116
 Qy 250 TyrMetGlyGlyThrSerMetAlaThrProIleValAlaGly---AsnValAlaGlnLeu 268
 Db 1117 AAGGCTCTGGAACCCAGCATGGCCACCGCGCGCTTTCGGCGGTGGCGCTCATCCTC 1176
 Qy 269 ArgGluHisPheValLysAsnArgGlyIleThrProLys-----ProSerLeuLeuLys 286
 Db 1177 CAGGCCAC-----CGAGCTGGACCCCGCAAGCAAGTGAAG 1212
 Qy 287 AlaAlaLeuIleAlaGlyAla-----AlaAspIleGlyLeu 298
 Db 1213 ACCGCCCTCATCGAGACCGCGCATAGTCGCCCAAGGAGATAGCGACATCGCTAC 1272
 Qy 299 GlyTyrrProAsnGlyAsnGlnGlyTyrrGlyArgValThrLeuAspLysSerLeu----- 316
 Db 1273 GGTGCG-----GGTAGGTGAACGCTTACAAGGCCATCAAGTAC 1311
 Qy 317 ---AsnValAlaTyrrValAsnGluSerSerSerLeuSerThrSerGlnLysAlaThrTyrr 335
 Db 1312 GACGACTACGCGCAAGCTCACCTTCACCGGCTCCGTCGCGCACAGGGAAGCGCCACCAC 1371
 Qy 336 SerPheThrAlaThrAlaGlyLysProLeuLysIleSerLeuValTrpSerAspAlaPro 355
 Db 1372 ACCTTCGACGTCAGCGCGCCACCTTCGTGACCCGCCACCTCTACTGGGAC----- 1422
 Qy 356 AlaSerThrThrAlaSerValThrLeuValAsnAspLeuLeuValIleThrAlaPro 375
 Db 1423 -----ACGGGCTCGAGCGACATCGACCTCTACTCTAGACCCC 1461
 Qy 376 AsnGlyThrGlnTyrrValGlyAsnAspPheThrSerProTyrrAsnAspAsnTrpAspGly 395
 Db 1462 AACGGGAACGAG---GTTGACTACTCTACACCGCCCTACTAC----- 1500
 Qy 396 ArgAsnAsnValGluAsnValPheIleAsnAlaProGlnSerGlyThrTyrrThrIleGlu 415
 Db 1501 -----GGCTTCGAGAGGTGGGTACTACACCCGCGCGGAACCTGACCGGTCAAG 1554
 Qy 416 ValGlnAlaTyrrAsnValProValGlyProGlnThrPheSerLeuAlaIleValAsn 434
 Db 1555 GTCGTCAGCTACAAG-----GGCGCGCGCAACTACCAGGTCGACGTCGTGCTCAGC 1602

| | | | |
|--|---|---|------|
| QY | 416 | VaIGlnAlaTyRAsnValProValGlyProGlnThrPheSerLeuAlaIleValAsn | 434 |
| Db | 1555 | GTCTGAGCTACAAG-----GGCGGGCGGAAGTACCAGGTGCGTCGTCAGC | 1602 |
| RESULT 8 | | | |
| AA05920 | | | |
| ID | AA05920 | standard; DNA; 1236 BP. | |
| XX | AA05920; | | |
| DT | 06-MAY-1999 | (first entry) | |
| XX | Hyperthermostable protease fragment encoding DNA. | | |
| XX | Hyperthermostable; protease; thermophilic; bacterium; subtilisin; | | |
| KW | additive; drug; washing agent; foodstuff; chemical synthesis; ds. | | |
| OS | Pyrococcus furiosus. | | |
| XX | Pyrococcus furiosus. | | |
| FN | W09856926-A1. | | |
| XX | 17-DEC-1998. | | |
| XX | 04-JUN-1998; | 98WO-JP002465. | |
| PF | 10-JUN-1997; | 97JP-00151969. | |
| PR | (TAKI) TAKARA SHUZO CO LTD. | | |
| XX | Takakura H, Morishita M, Shimojo T, Asada K, Kato I; | | |
| XX | WPI; 1999-080907/07. | | |
| DR | P-PSDB; AA094836. | | |
| XX | Recombinant hyperthermostable protease from Pyrococcus furiosus - and | | |
| PT | gene encoding it, for large scale production of the protease for | | |
| PT | industrial use. | | |
| XX | Claim 6; Page 37-38; 82pp; Japanese. | | |
| XX | The invention relates to a hyperthermostable protease derived from a | | |
| CC | thermophilic bacterium (especially Pyrococcus furiosus). The protease has | | |
| CC | working temperature 40-110 deg.C (optimum 80-95 deg.C), working pH 5-10 | | |
| CC | (optimum 6-8), and retains more than 90% of its activity after 8 hours at | | |
| CC | 95 deg.C. The invention also provides gene sequences encoding a | | |
| CC | polypeptide of formula SIG-Ala-Gly-Gly-Asn-PRO, where SIG is a signal | | |
| CC | peptide from subtilisin, and PRO is the above protease. Host cells | | |
| CC | (especially Bacillus strains) transformed with vectors comprising the | | |
| CC | genes are used for the recombinant production of the protease. The | | |
| CC | hyperthermostable protease which can be prepared in quantity suitable for | | |
| CC | industrial use, can be used as an additive for drugs, washing agents and | | |
| CC | foodstuffs and for chemical synthesis | | |
| XX | Sequence 1236 BP; 368 A; 251 C; 309 G; 308 T; 0 U; 0 Other; | | |
| SQ | | | |
| Alignment Scores: | | | |
| Pred. No.: | 1.59e-22 | Length: | 1236 |
| Score: | 411.50 | Matches: | 140 |
| Percent Similarity: | 42.55% | Conservative: | 60 |
| Best Local Similarity: | 29.79% | Mismatches: | 148 |
| Query Match: | 18.31% | Indels: | 122 |
| DB: | 2 | Gaps: | 20 |
| US-09-985-689A-1-COPY (1-434) x AA05920 (1-1236) | | | |
| QY | 12 | ValAlaGlnSerSerTyRgLy-LeuTyr-----GlyGlnGlyG1 | 24 |
| Db | 24 | GTCTGCAGCTCAAGTTATGGCACTTACGTTGGAACTTGGGATATGATGTTCTGGAAT | 83 |
| QY | 24 | nileValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerMetHisG1 | 44 |
| Db | 84 | CACAATAGGAATAATTGACACTGGAATTGAC-----GCTTCTCATCC | 125 |

| | | | |
|----|------|--|------|
| QY | 44 | uAlaPheArgGlyLysIleThrAlaLeuTyRAlaLeuGlyArgThrAsnAsnAlaAsn-- | 63 |
| Db | 126 | AGATCTCCAGAAAGTA-----ATTGGGTGGGTAGATTTTGTCAATGG | 170 |
| QY | 64 | -----AspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuG1 | 78 |
| Db | 171 | TAGGAGTTATCCATACGATGACATGACATGGAACTCAGTCTCAATAGCAGCTGG | 230 |
| QY | 78 | yAsnGlySerThrAsn-----LysGlyMetAlaProGlnAlaAsnLeuValPh | 94 |
| Db | 231 | TACTGGAGCAGCAAGTAATGCAAGTACAAGGGAATGGCTCCAGGAGTAACTTGGGGG | 290 |
| QY | 94 | eGlnSerIleMet-----AspSerGlyGlyGlyLeuGlyGlyLeuProSerAsnLeuG1 | 112 |
| Db | 291 | AATTAAGTTCTAGTCCGAGTCTTGGAGCATATCTACTATAATTAAGGAGTTGA | 350 |
| QY | 112 | nThrLeuPheSerGlnAlaTyRSerAlaGlyAlaArgIleHisThrAsnSerTrpGlyAl | 132 |
| Db | 351 | GTGGGCCGCTTGATAACAAAGATAAGTACGGAATTAAAGTCAATTAATCTTCTCTGTTTC | 410 |
| QY | 132 | a-----AlaValAsnGlyAlaTyRThr | 140 |
| Db | 411 | AAGCCAGAGCTCAGATGGTACTGACGCTTAAGTCAGGCTGTTAATCGACGGTGGATGC | 470 |
| QY | 140 | rAspSerArgAsnValAspAspTyRValArgLysAsnAspMetThrIleLeuPheAlaAl | 160 |
| Db | 471 | T-----GGATTAGTTGTGTGTTGCCGC | 494 |
| QY | 160 | aGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaI1 | 180 |
| Db | 495 | TGGAACACAGTGGACCTTACAAGTATACAATCGTTCTCCAGCAGCTGCCAGCAAGTTAT | 554 |
| QY | 180 | eThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyRAlaAspAsnIleAs | 200 |
| Db | 555 | TACAGTTGGAGCC-----GTTGACAGTATGA | 581 |
| QY | 200 | nHisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVa | 220 |
| Db | 582 | TGTTATAACAAGCTTCTCAAGCAGAGGGCCAACTTACAGACAGCTCTGGGACATCAATGGCACTCTCA | 752 |
| QY | 240 | eTrpAlaAsnHisAspSerLysTyRAlaTyRMetGlyGlyThrSerMetAlaThrProI1 | 260 |
| Db | 693 | GGGTCAACCAATTAATGACTATTACACAGCAGCTCTCTGGGACATCAATGGCACTCTCA | 752 |
| QY | 260 | eValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPr | 280 |
| Db | 753 | CGTAGCTGGTATTGAGCCCTCTTGTCTCAA-----GCACACCC | 791 |
| QY | 280 | oLys-----ProSerLeuLysAlaAlaLeuIleAlaGlyAla----- | 293 |
| Db | 792 | GAGCTGGACTCCAGACAAAGTAAACAGCCCTCATAGAACTGCTGATATCGTAAAGCC | 851 |
| QY | 294 | -----AlaAspIleGlyLeuGlyTyRProAsnGlyAsnGlnGlyTrpGlyArgVa | 310 |
| Db | 852 | AGATGAAATAGCCGATATAGCTACGGTGCA-----GGTAGGGT | 890 |
| QY | 310 | lThrLeuAspLysSerLeuAsnValAlaTyRValAsnGluSerSerSerLeuSerThrSe | 330 |
| Db | 891 | TAATGCATACAGGCTATAAC-----TACGATACTATGCAGAGCTAGTGTTCACCTGG | 944 |
| QY | 330 | rGlnLysAla-----ThrTyRSerPheThrAlaThrAlaGlyLysProLe | 345 |
| Db | 945 | ATATGTGGCAACAAAGGCAGCCAACTCACAGTTCGTTATTAGCGGAGCTTCGTTCTGT | 1004 |
| QY | 345 | uLysIleSerLeuValTrpSerAspAlaProAlaSerThrThrAlaSerValThrLeuVa | 365 |
| Db | 1005 | ACTGCCCATTTACTGGACATGCCAAT-----GCTTCTCATCC | 1035 |
| QY | 365 | lAsnAspLeuAsnLeuValIleThrAlaProAsnGlyThrGlnTyRValGlyAsnAspPh | 385 |

```

Db 1036 -AGGACCTTGATCTTACCCTACGATCCCAATGGAACCCAG---GTGACTACTCTTA 1091
QY 385 eThrSerProTyrAsnAspAsnTrpAspGlyArgAsnAsnValGluAsnValPheIleAs 405
Db 1092 CACCGCCTACTAT-----GGATTCGAAAGGTTGTTATTA 1127
QY 405 nAlaProGlnSerGlyThrThrIleGluValGlnAlaTyrAsnValProValGlyPr 425
Db 1128 CAAACCAACTGATGGAACATGACAAATTAAGTTCTTAAGCTACAGC-----GGAAG 1178
QY 425 oGlnThrPheSerLeuAlaIleValAsn 434
Db 1179 TGCAAACTATCAAGTAGATGTGTAAGT 1206

RESULT 9
ART85668
ID AAT85668 standard; DNA; 1566 BP.
XX
XX AAT85668;
AC AC
DT 17-OCT-2003 (revised)
DT 20-APR-1998 (first entry)
XX
DE Pyrococcus furiosus protease coding sequence.
XX
XX Protease; research reagent; thermal stability; pyrococcus furiosus; ss.
XX
XX Pyrococcus furiosus; DSM-3638.
XX
XX Key Location/Qualifiers
XX CDS 1..1566
XX FT /cag= a
XX FT /transl_except= (pos: 1282..1284, aa: Xaa)
XX FT /note= "Xaa= Gly, Val"
XX
XX WO921823-A1.
XX
XX 19-JUN-1997.
XX
XX 07-NOV-1996; 96WO-JP003253.
XX
XX 12-DEC-1995; 95JP-00323285.
XX
XX (TAKI ) TAKARA SHUZO CO LTD.
XX
XX Takakura H, Morishita M, Yamamoto K, Mitta M, Asada K;
XX Tsunabawa S, Kato I;
XX WPI; 1997-332794/30.
XX P-PSDB; AAW24122.
XX
XX Protease(s) and genes encoding them obtained from Thermococcus and
XX Pyrococcus strains - have extremely high thermal stability and are useful
XX industrially and as research reagents.
XX
XX Claim 7; Page 90-91; 159pp; Japanese.
XX
XX This sequence represents the coding sequence for the protease from
XX Pyrococcus furiosus DSM-3638. This sequence encodes a protease of the
XX invention. The proteases of the invention have extremely high thermal
XX stability. The proteases can be used as research reagents, and
XX industrially in the food, drug and chemical industries. (Updated on 17-
XX OCT-2003 to standardise OS field)
XX
XX Sequence 1566 BP; 467 A; 342 C; 372 G; 384 T; 0 U; 1 Other;

Alignment Scores:
Pred. No.: 2.11e-22 Length: 1566
Score: 411.50 Matches: 140
Percent Similarity: 42.55% Conservative: 60
Best Local Similarity: 29.79% Mismatches: 148
Query Match: 18.31% Indels: 122

```

```

DB: 2 Gaps: 20
US-09-985-689a-1-COPY (1-434) x AAT85668 (1-1566)
QY 12 ValAlaGlnSerSerTyrGly-LeuTyr-----GlyGlnGly 24
Db 24 GTTCGAGCTCAAGTTATGCGCACTTACGTTTGGAACTGATATGATGGTTCTGGAAT 83
QY 24 nIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerSerMetHisG1 44
Db 84 CACAATAGGAATAATTGACACTGGAATTGAC-----GCTTCTCATCC 125
QY 44 uAlaPheArgGlyValIleThrAlaLeuTyrAlaLeuGlyArgThrAsnAsnAlaAsn-- 63
Db 126 AGATCTCCAAGGAAAGTA-----ATTGGTGGGTAGATTTGTCAATGG 170
QY 64 -----AspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuG1 78
Db 171 TAGGAGTTATCCATACGATGACCATGACATGCACTCATGTAGCTTCAATAGCAGCTGG 230
QY 78 yAsnGlySerThrAsn-----LysGlyMetAlaProGlnAlaAsnLeuValPh 94
Db 231 TACTGGAGCAGCAAGTAATGGCAAGTACAGGAATGGCTCCAGGAGCTAAGCTGGCGGG 290
QY 94 eGlnSerIleMet-----AspSerGlyGlyLeuGlyGlyLeuProSerAsnLeuG1 112
Db 291 AATTAGGTTCTAGGTGCGGATGGTTCTGGAAGCATATCTACTATAATTAAGGAGCTGA 350
QY 112 nThrLeuPheSerGlnAlaTyrSerAlaGlyAlaArgIleHisThrAsnSerTrpGlyAl 132
Db 351 GTGGCGCGTTGATACAAAGATAAGTACGGAATTAAGGTCATTAATCTTCTCTTCTGTT 410
QY 132 a-----AlaValAsnGlyValaTyrThrTh 140
Db 411 AAGCCAGAGCTCAGATGGTACTGACGCTCTTAAGTCAGGCTGTTAATGACGCGTGGATGC 470
QY 140 rAspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAla 160
Db 471 T-----GGATTAGTTGTTGTGTTGCCGC 494
QY 160 aGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAla 180
Db 495 TGGAAACAGTGGACCTTACCAAGTATACATCGGTTCTCCAGCAGCTGCAAGCAAGTTAT 554
QY 180 eThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAs 200
Db 555 TACAGTTGGAGCC-----GTTGACAGATATGA 581
QY 200 nHisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVa 220
Db 582 TGTTATTAACAAGCTTCTCAAGCAGAGGGCCCACTGCAGACGGCAGGCTTAAGCTGAGGT 641
QY 220 lMetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPh 240
Db 642 TGTGTCCTCAGAAACTGGATAATTGCTGCCAGAGCAAGT-----GGAACTAGCAT 692
QY 240 eTrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrPro 260
Db 693 GGGTCAACCAATTATGACTATTATACAGCAGGCTCCTGGGACATCATGCGCACTCTCA 752
QY 260 eValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPr 280
Db 753 CGTAGCTGGTATTGACGCGCTCTTGTCTCAA-----GCACACCC 791
QY 280 oLys-----ProSerLeuLeuLysAlaAlaIleAlaGlyAla----- 293
Db 792 GAGCTGGACTCCAGCAAAAGTAAACAGCCCTCATAGAACTGCTATATATCGTAAGGCC 851
QY 294 -----AlaAspIleGlyLeuGlyTyrProAsnGlyAsnGlnGlyTrpGlyArgVa 310
Db 852 AGATGAATAGCCGATATAGCTACGGTGCA-----GGTAGGGT 890
QY 310 lThrLeuAspLysSerLeuAsnValAlaTyrValAsnGluSerSerSerLeuSerThrSe 330

```

Db 891 TAATGCATACAGGCTATAAC-----TACGATAACTATGCAAAAGCTAGTGTCTACCTGG 944
Qy 330 rGlnLysAla-----ThrTyrSerPheThrAlaThrAlaGlyLysProLe 345
Db 945 ATATGTTGCCAACAAAGGACCAAACTCACAGTTGGTTATTAGCGGAGCTTCGTTCCT 1004
Qy 345 uLysIleSerLeuValTrpSerAspAlaProAlaSerThrThrAlaSerValThrLeuVa 365
Db 1005 AACTGCCACATTACTGGGACCAATGCCAAT----- 1035
Qy 365 lAsnAspLeuAsnLeuValIleThrAlaProAsnGlyThrGlnTyrValGlyAsnAspPh 385
Db 1036 -AGGACCTTGATCTTACTCTACAGTCCCAATGGAAACAG--GTTGACTACTCTTA 1091
Qy 385 eThrSerProTyrAsnAspAsnTrpAspGlyA-gAsnAsnValGluAsnValPheIleAs 405
Db 1092 CACCGCCTACTAT-----GGATTGCAAAAGGTTGGTTATTA 1127
Qy 405 nAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsnValProValGlyPr 425
Db 1128 CAACCCAACTGATGGAACTAGGCAATTAAGTTGTAAGCTACAGC-----GGAAG 1178
Qy 425 oGlnThrPheSerLeuAlaIleValAsn 434
Db 1179 TGCAAACTATCAAGTAGATGTGGTAAGT 1206

RESULT 10

AAT85695
ID AAT85695 standard; DNA; 1962 BP.
XX AC AAT85695;
XX DT 17-OCT-2003 (revised)
XX DT 20-APR-1998 (first entry)
XX DE Pyrococcus furiosus PFUS protease coding sequence.
XX KW Protease; research reagent; thermal stability; pyrococcus furiosus; ss.
XX OS Pyrococcus furiosus; DSM-3638.
XX PN W09721823-A1.
XX PD 19-JUN-1997.
XX PF 07-NOV-1996; 96WO-JP003253.
XX PR 12-DEC-1995; 95JP-00323285.
XX PA (TAKI) TAKARA SHUZO CO LTD.
XX PI Takakura H, Moxishita M, Yamamoto K, Mitta M, Asada K;
XX PI Teunaeawa S, Kato I;
XX PI WPI; 1997-332794/30.
XX DR P-FSDB; AAW24129.
XX PT Protease(s) and genes encoding them obtained from Thermococcus and
XX PT Pyrococcus strains - have extremely high thermal stability and are useful
XX PT industrially and as research reagents.
XX PS Disclosure; Page 123-125; 159pp; Japanese.
XX CC This sequence represents the coding sequence for the protease from
XX CC Pyrococcus furiosus DSM-3638. This sequence encodes a protease of the
XX CC invention. The proteases of the invention have extremely high thermal
XX CC stability. The proteases can be used as research reagents, and
XX CC industrially in the food, drug and chemical industries. (Updated on 17-
XX CC OCT-2003 to standardise OS field)
XX SQ Sequence 1962 BP; 602 A; 399 C; 471 G; 490 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.: 2,76e-22 Length: 1962
Score: 411.50 Matches: 140
Percent Similarity: 42.55% Conservative: 60
Best Local Similarity: 29.79% Mismatches: 148
Query Match: 18.31% Indels: 122
DB: 2 Gaps: 20
US-09-985-689A-1-COPY (1-434) x AAT85695 (1-1962)
Qy 12 ValAlaGlnSerSerTyrGly-LeuTyr-----GlyGlnGlyG1 24
Db 420 GTCTGCAGCTCAAGTTATGGCAACTTACGTTTGGAACTTGGGATATGATGTTCTGGAAT 479
Qy 24 nIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisG1 44
Db 480 CACAAATAGGAATAATTGACACTGGAAATGAC-----GCTTCTCATCC 521
Qy 44 uAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsnAsnAlaAen-- 63
Db 522 AGATCTCCCAAGGAAAGTA-----ATTGGGTGGGTAGATTTTGTCAATGG 566
Qy 64 -----AspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuG1 78
Db 567 TAGGAGTTATCCATACGATGACCATGGACATGAGTCTATGCTTCAATAGCAGCTGG 626
Qy 78 yAsnGlySerThrAsn-----LysGlyMetAlaProGlnAlaAsnLeuValPh 94
Db 627 TACTGGACGACCAAGTAATGCAAGTACAGGAATGCTCCAGGAGCTAAGCTGGCGG 686
Qy 94 eGlnSerIleMet-----AspSerGlyGlyGlyLeuGlyGlyLeuProSerAsnLeuG1 112
Db 687 AATTAAGGTTCTAGGTGCGCATGTTCTGGAAGCATATCTACTATAATTAAGGAGTTGA 746
Qy 112 nThrLeuPheSerGlnAlaTyrSerAlaGlyAlaValAileHisThrAsnSerTrpGlyAl 132
Db 747 GTGGCCCGTTGATACAAAGATAGTACGGAATTAAGTCAATTAATCTTCTTCTGTTTC 806
Qy 132 a-----AlaValAsnGlyAlaTyrThrTh 140
Db 807 AAGCCAGAGCTCAGATGCTACTGACGCTCTAAGTCAGGCTTTAATGCGAGCTGGGATG 866
Qy 140 rAspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAl 160
Db 867 T-----GGATTAGTTGTTGGTTGCCG 890
Qy 160 aGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaI1 180
Db 891 TGGAAACAGTGGACCTCAAGTATACAACTCGGTCTCCAGCAGCTGCAAGCAAGTTAT 950
Qy 180 eThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAs 200
Db 951 TACAGTTGGAGCC-----GTTGCAAGTATGA 977
Qy 200 nHisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVa 220
Db 978 TGTATTAACAAGCTTCTCAAGCAGAGGCGCAACTGCAGCGGAGCTTAAGCTGAGGT 1037
Qy 220 lMetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPh 240
Db 1038 TGTTCCTCCAGAACTCGAATAATTGCTGCCAGCAAGT-----GGAATCTAGAT 1088
Qy 240 eTrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyThrSerMetAlaThrProI1 260
Db 1089 GGGTCAACCAATTATGACTATTACAGCAGAGCTCTCTGGACATCATATGGCACTCTCTCA 1148
Qy 260 eValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPr 280
Db 1149 CGTAGCTGTATTCGACGCCCTCTTGCTCCAA-----GCACACCC 1187
Qy 280 oLys-----ProSerLeuLeuLysAlaAlaLeuIleAlaGlyAla----- 293
Db 1188 GAGCTGGACTCCAGCAAAAGTAAACAGCCCTCATAGAACTGCTGATATCGTAAAGCC 1247

QY 294 -----AlaAspIleGlyLeuGlyTyrProAsnGlyAsnGlnGlyTyrGlyArgVa 310
 Db 1248 AGATGAATAGCGGATATAGCTACGGTGCA-----GGTAGGGT 1286
 QY 310 lThrLeuAspLysSerLeuAsnValAlaTyrValAsnGluSerSerLeuSerThrSe 330
 Db 1287 TAATGCATCAAGGCTATAAAC-----TACGATAACTATGCAAGAGTAGTTCACCTGG 1340
 QY 330 rGlnLysAla-----ThrTyrSerPheThrAlaThrAlaGlyLysProLe 345
 Db 1341 ATATGTTGCCAACAAAGGCGAGCAACTCACCAGTTGGTTATTAGCGGAGTTCGTTGCT 1400
 QY 345 uIlyleSerLeuValTyrSerAspAlaProAlaSerThrThrAlaSerValThrLeuVa 365
 Db 1401 AACTGCCACATTATATCGGACAAATGCCAAT----- 1431
 QY 365 lAsnAspLeuAsnLeuValIleThrAlaProAsnGlyThrGlnTyrValGlyAsnAspPh 385
 Db 1432 -ACGGACCTTGATCTTACTCTACGATCCCAATGGAAACCAAG---GTTGACTACTCTTA 1487
 QY 385 eThrSerProTyrAsnAspAsnTrpAspGlyArgAsnAsnValGluAsnValPheIleAs 405
 Db 1488 CACCGGCTACTAT-----GGATTGGAAGGTTGGTTATTA 1523
 QY 405 nAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsnValProValGlyPr 425
 Db 1524 CAACCCAACTGATGGACATGGCAATTAAGTTGTACTACAGC-----GGAAG 1574
 QY 425 oGlnThrPheSerLeuAlaIleValAsn 434
 Db 1575 TGCAAACTATCAAGTAGATGTGTAAGT 1602

RESULT 11

AAX05929
 ID RAX05929 standard; DNA; 1962 BP.

XX AAX05929;
 AC AC
 XX XX
 DT 06-MAY-1999 (first entry)
 XX XX

DE Hyperthermostable protease encoding DNA.

XX Hyperthermostable; protease; thermophilic; bacterium; subtilisin;
 XX additive; drug; washing agent; foodstuff; chemical synthesis; ds.

OS Pyrococcus furiosus.

XX W09856926-A1.

XX 17-DEC-1998.

XX 04-JUN-1998; 98WO-JP002465.

XX 10-JUN-1997; 97JP-00151969.

XX (TAKI) TAKARA SHUZO CO LTD.

XX Takakura H, Morishita M, Shimojo T, Asada K, Kato I;

XX WPI: 1999-080907/07.

XX P-PSDB; AAW94841.

PT Recombinant hyperthermostable protease from *Pyrococcus furiosus* - and
 PT gene encoding it, for large scale production of the protease for
 PT industrial use.

PS Disclosure; Page 59-60; 82pp; Japanese.

CC The invention relates to a hyperthermostable protease derived from a
 CC thermophilic bacterium (especially *Pyrococcus furiosus*). The protease has
 CC working temperature 40-110 deg.C (optimum 80-95 deg.C), working pH 5-10
 CC (optimum 6-8), and retains more than 90% of its activity after 8 hours at

CC 95 deg.C. The invention also provides gene sequences encoding a
 CC polypeptide of formula SIG-Ala-Gly-Gly-Asn-PRO, where SIG is a signal
 CC peptide from subtilisin, and PRO is the above protease. Host cells
 CC (especially *Bacillus* strains) transformed with vectors comprising the
 CC genes are used for the recombinant production of the protease. The
 CC hyperthermostable protease which can be prepared in quantity suitable for
 CC industrial use, can be used as an additive for drugs, washing agents and
 CC foodstuffs and for chemical synthesis

XX SQ Sequence 1962 BP; 602 A; 399 C; 471 G; 490 T; 0 U; 0 Other;

Alignment Scores: 2,76e-22 Length: 1962
 Pred. No.: 411.50 Matches: 140
 Score: 411.50 Conserved: 60
 Percent Similarity: 42.55%
 Best Local Similarity: 29.79%
 Query Match: 18.31%
 Indels: 122
 Gaps: 20

US-09-985-689A-1-COPY (1-434) x AAX05929 (1-1962)

QY 12 ValAlaGlnSerSerTyrGly-LeuTyr-----GlyGlnGlyG1 24
 Db 420 GTCTGCAGCTCAAGTTATGGCAACTTACGTTTGGAACTTCGGATATGATGTTCTGGAT 479
 QY 24 nIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisG1 44
 Db 480 CACAATAGGAATAATTGCACACTGGAATTGAC-----GCTTCTCATCC 521
 QY 44 uAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsnAsnAlaAsn-- 63
 Db 522 AGATCTCCNAGGAAGTA-----ATTGGTGGGTAGATTCTTCATATGG 566
 QY 64 -----AspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuG1 78
 Db 567 TAGGATTATCCATACCATGACATGCATGACATGACATGATAGTTCATATAGCAGCTGG 626
 QY 78 YAsnGlySerThrAsn-----LysGlyMetAlaProGlnAlaAsnLeuValPh 94
 Db 627 TACTGGACGACCAAGTAATGCAAGTACAAAGGGAATGCTCCAGGAGCTTAAGCTGGCGGG 686
 QY 94 eGlnSerIleMet-----AspSerGlyGlyGlyLeuGlyGlyLeuProSerAsnLeuG1 112
 Db 687 AATTAAAGTTCTAGTGCCGATGTTCTGGAGCAATATCTACTATAATTAAGGAGTTGA 746
 QY 112 nThrLeuPheSerGlnAlaTyrSerAlaGlyAlaArgIleHisThrAsnSerTrpGlyAl 132
 Db 747 GTGGGCGCTTCATATAACAAAGATAAGTACGGAATTAAAGTTCATTATCTTCTCTGTTTC 806
 QY 132 a-----AlaValAsnGlyAlaTyrThrTh 140
 Db 807 AAGCCAGAGCTCAGATGGTACTGACGCTCTAAGTCAGGCTGTTAATGACAGGTGGATGC 866
 QY 140 rAspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaI 160
 Db 867 T-----GGATTAGTTGTTGTTGTTGCGGC 890
 QY 160 aGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaI 180
 Db 891 TGGAAACAGTGGACCTTACAAATGATACATCGTTCTCCAGCAGCTGCAACCAAGTTAT 950
 QY 180 eThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAs 200
 Db 951 TACAGTTGGAGCC-----GTTGACAAAGTATGA 977
 QY 200 rHisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVa 220
 Db 978 TGTATTATAACAGCTTCTCAACGACAGGCGCAACTGCGACGCGGAGGCTTAAGCTGAGGT 1037
 QY 220 lMetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerPh 240
 Db 1038 TGTGTCTCCGGAACACTGATAATTGCTGCCAGAGCAAGT-----GGAAGTATGAT 1088

QY 240 eTpaAaSHiAspSerLysTyrAlaTyrMetGlyThrSerMetAlaThrProI 260
 Db 1089 GGFTCAACCAATTAATGACTATTACACAGCAGCTCCTGGGACATCAATGGCACTCTCA 1148
 QY 260 eValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyLleThrP 280
 Db 1149 CGTAGCTGGATTGACGCCCTCTGTCTCAA-----GCACACCC 1187
 QY 280 oLys-----ProSerLeuLeuLysAlaAlaLeuLleAlaGlyAla----- 293
 Db 1188 GAGCTGGACTCCAGACAAAGTAAACAGCCCTCATAGAACTGCTGATATCGTAAAGCC 1247
 QY 294 -----AlaAspIleGlyLeuGlyTyrProAsnGlyAsnGlnGlyTrpGlyArgVa 310
 Db 1248 AGATGAATACCGCATATACCTACGCTGCA-----GGTAGGTT 1286
 QY 310 lThrLeuAspLysSerLeuAsnValAlaTyrValAsnGluSerSerSerLeuSerThrSe 330
 Db 1287 TAATGCATACAGCTATAAAC-----TACGATACTATGCAAGCTAGTGTCTACTGG 1340
 QY 330 rGlnLysAla-----ThrTyrSerPheThrAlaThrAlaGlyLysProLe 345
 Db 1341 ATATGTTGCCAACAAAGGCGCAAACTCACCAGTTTCGTTATTAGCGGAGCTTCGTTCTG 1400
 QY 345 uLysIleSerLeuValTrpSerAspAlaProAlaSerThrThrAlaSerValThrLeuVa 365
 Db 1401 AACTGCCACATTATCTGGGACATGCAAT----- 1431
 QY 365 lAsnAspLeuAsnLeuValIleThrAlaProAsnGlyThrGlnTyrValGlyAsnAspPh 385
 Db 1432 -AGCGACTTGATCTTTACCTCTACGATCCCAATGGAAACGAG---GTTGACTACTCTTA 1487
 QY 385 eThrSerProTyrAsnAspAsnTrpAspGlyArgAsnValGluAsnValPheIleAs 405
 Db 1488 CACCGCTTACTAT-----GGATTGCAAAAGGTTGGTTATTA 1523
 QY 405 nAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsnValProValGlyPr 425
 Db 1524 CAACCACTGATGGACATGGACAAATTAGGTTGTAAAGCTACAGC-----GGAAG 1574
 QY 425 oGlnThrPheSerLeuAlaIleValAsn 434
 Db 1575 TGCAAACTCAAGTAGATGGTAAAGT 1602

RESULT 12
 AAT85669
 ID AAT85669 standard; DNA; 1977 BP.
 XX AC AAT85669;
 XX DT 20-APR-1998 (first entry)
 XX DE Protease coding sequence.
 XX KW Protease; research reagent; thermal stability; ss.
 XX OS Synthetic.
 XX PN WO9721823-A1.
 XX PD 19-JUN-1997.
 XX PF 07-NOV-1996; 96WO-JP003253.
 XX PR 12-DEC-1995; 95JP-00323285.
 XX PA (TAXI) TAKARA SHUZO CO LTD.
 XX PI Takakura H, Morishita M, Yamamoto K, Mitta M, Asada K;
 XX PI Tsurasawa S, Kato I;
 XX DR WPI; 1997-332794/30.
 XX DR P-PSDB; AAM24123.

XX Protease(s) and genes encoding them obtained from Thermococcus and
 PT Pyrococcus strains - have extremely high thermal stability and are useful
 PT industrially and as research reagents.
 XX Claim 11; Page 95-97; 159pp; Japanese.
 XX This sequence represents the coding sequence for a protease of the
 CC invention. The proteases of the invention have extremely high thermal
 CC stability. The proteases can be used as research reagents, and
 CC industrially in the food, drug and chemical industries
 XX Sequence 1977 BP; 527 A; 562 C; 513 G; 375 T; 0 U; 0 Other;

Alignment Scores:

Pred. No.: 2,78e-21 Length: 1977
 Score: 398.50 Matches: 133
 Percent Similarity: 43.07% Conservative: 69
 Best Local Similarity: 28.36% Mismatches: 147
 Query Match: 17.73% Indels: 120
 DB: 2 Gaps: 20

US-09-985-689A-1-COPY (1-434) x AAT85669 (1-1977)

QY 12 ValAlaGlnSerSerTyrGly-LeuTyr-----GlyGlnGly 24
 Db 420 GTCTGCAGCTCAAGTTATGGCACTTACGTTTGGAACTTGGATATGATGTTCTTGGAA 479
 QY 24 nIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisG 44
 Db 480 CACAATAGGAATAATTGCACTGGAATTGAC-----GCTTCTCATCC 521
 QY 44 uAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsnAsnAlaAsn-- 63
 Db 522 AGATCTCCAAGAAAGTA-----ATTGGTGGTAGATTTTGTCAATGG 566
 QY 64 -----AspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuG 78
 Db 567 TAGGAGTTATCCATACGATGACCATGCGATGCACTCATGTAGTTCATAGCAGCTGG 626
 QY 78 vAsnGlySerThrAsn-----LysGlyMetAlaProGlnAlaAsnLeuValPh 94
 Db 627 TACTGGACGCAAGTATGCAAGTACAGGAATGCTCCAGGAGCTAAGCTGGCGGG 686
 QY 94 eGlnSerIleMet-----AspSerGlyGlyGlyLeuGlyGlyLeuProSerAsnLeuG 112
 Db 687 AATTAAGGTTCTAGGTGCGATGTTCTGGAAGCATATCTACTATAATAAAGGAGTTGA 746
 QY 112 nThrLeuPheSerGlnAlaTyrSerAlaGlyAlaArgIle----- 125
 Db 747 GTGGCCGTTGATACAAAGATAAGTACGAATTAGGTCAATATCTTCTTGGTTC 806
 QY 126 -----HisThrAsnSerTrpGlyAlaAlaValAsnGlyAlaTyrThrTh 140
 Db 807 AAGCCAGAGCTCCGACGGAACCGACTCCCTCAGTCAGGCCGCTCAACACCGCTGGAGCG 866
 QY 140 rAspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAl 160
 Db 867 C-----GGTAGTAGTCTGCTCGCCGC 890
 QY 160 aGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaI 180
 Db 891 CGGCAACAGCGGGCGCAACACCTACACCTCGCTCCGCCGCCGCGGAGCAAGTCA 950
 QY 180 eThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAs 200
 Db 951 AACCGTCGGTGCA-----GTTGACAGCAACGA 977
 QY 200 nHisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVa 220
 Db 978 CACATCGCGAGCTTCTCCAGGAGGGACCGACCGGACGAGGCTCAAGCCGAGCT 1037
 QY 220 lMetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPh 240

```

Db      1038  CGTCGCCCGCGGCTTGACATCATACCCCGCGCCGACG-----GGAACGAGCAT 1088
Qy      240  eTPlaaAsnHieAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIle 260
Db      1089  GGCCACCCCGGATACGACTACTACCAAGGCGCTCTGGACACGAGCATGGCCACCCGCA 1148
Qy      260  eValAlaGly---AsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyLeu 279
Db      1149  CGTTTCGGGCGTTGGCGCGCTCATCTCCAGGCCAC----- 1185
Qy      279  rProLys-----ProSerLeuLeuLysAlaAlaLeuIleAlaGlyAla----- 293
Db      1186  -CGAGCTGACCCCGGACAGGTGAAGACCGCGCTCATCGAGACCGCGCATATGCG 1244
Qy      294  -----AlaAspIleGlyLeuGlyTyrProAsnGlyAsnGlnGlyTrpGlyAr 309
Db      1245  CCCCAAGGAGATACCGGACATCCCTACGCTGCG-----GGTAG 1283
Qy      309  gValThrLeuAspLysSerLeu-----AsnValAlaTyrValAsnGluSerSe 326
Db      1284  GGTGAAGCTCTACAGGCCATCATAGTACGACGACTACGCCAAGCTCACCTTCCGGCTC 1343
Qy      326  rLeuSerThrSerGlnLysAlaThrTyrSerPheThrAlaThrAlaGlyLysProLeu 346
Db      1344  CGTCGCCGACAGGAAGCGCCACCCACACCTTCGACGTCAGCGCGCCACCTTCGTGAC 1403
Qy      346  sIleSerLeuValTrpSerAspAlaProLaserThrThrAlaSerValThrLeuValAs 366
Db      1404  CGCCACCTCTACTGGAC-----ACGGGCTCGAG 1433
Qy      366  nAspLeuAsnLeuValIleThrAlaProAsnGlyThrGlnTyrValGlyAsnAspPhe 386
Db      1434  CGACATCGACCTCTACTCTCTACGCCCAACCGGAACGAG---GTTGACTACTCTCTAC 1490
Qy      386  rSerProTyrAsnAspAsnTrpAspGlyArgAsnAsnValGluAsnValPheIleAsn 406
Db      1491  CGCCTACTAC-----GGCTTCGAGAAGGTCCGGTACTACAA 1526
Qy      406  aProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsnValProValGlyPro 426
Db      1527  CCCGACCGCGGAACCTGGACGGTCAAGGTCTGTCAGCTACAAG-----GGCGCGC 1577
Qy      426  nThrPheSerLeuAlaIleValAsn 434
Db      1578  GAATACACGAGTGGACGCTGCTGACG 1602

RESULT 13
ABL54900
ID      ABL54900 standard; DNA; 2121 BP.
AC      ABL54900;
XX
XX
DT      11-SEP-2003 (revised)
DT      31-MAY-2002 (first entry)
XX
XX
DE      T. yonsei subtilisin-like serine protease coding sequence.
XX
XX
KW      Subtilisin-like serine protease; ss.
XX
XX
OS      Thermoanaerobacter yonseiensis.
XX
XX
FH      Key Location/Qualifiers
FT      CDS 142..1779
FT      /tag=a
FT      /product= "subtilisin-like serine protease"
XX
XX
KN      KR2000072141-A.
XX
XX
PD      05-DEC-2000.
XX
XX
PF      04-AUG-2000; 2000KR-00045411.
XX
XX

```

```

PR      04-AUG-2000; 2000KR-00045411.
XX
XX      (KIMY/) KIM Y S.
PI      Chang HJ, Kim DH, Byun YR, Kim YS;
XX
XX      WPI; 2001-298092/31.
XX      P-FSDB; ABB09483.
XX
XX      New DNA sequence of thermophilic protein decomposition enzyme and protein
XX      derived therefrom.
XX
XX      Claim 1; Page 6; 15pp; Korean.
XX
XX      This sequence represents the DNA encoding the Thermoanaerobacter yonsei
XX      subtilisin-like serine protease of the invention. (Updated on 11-SEP-2003
XX      to standardise OS field)
XX
XX      Sequence 2121 BP; 712 A; 410 C; 425 G; 554 T; 0 U; 20 Other;

Alignment Scores:
Pred. No.:      3,29e-19      Length:      2121
Score:          372.00      Matches:    130
Percent Similarity: 46.29%      Conservative: 51
Best Local Similarity: 33.25%      Mismatches: 114
Query Match:    16.56%      Indels:     96
DB:             4           Gaps:       19

US-09-985-689A-1-COPY (1-434) x ABL54900 (1-2121)

Qy      6      GlyIleValLysAlaAspValAlaGlnSerSerTyrGlyLeuTyrGlyGlnGlyGlnIle 25
Db      445  GGAATCAAAA-----GCACGAGTGATTTGGAGTCACAGGAAAAAATAAACA 495
Qy      26  ValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisGluAla 45
Db      496  ATAGCAATTATTGACACAGGTATAGCGGAATCACGTTGACCTCA----- 543
Qy      46  PheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsnAsnAlaAsn----- 63
Db      544  ---GGTGGAAAAATA-----ATAGGATGGAAGACTTTTATCAACACAAA 585
Qy      64  -----AspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsn 79
Db      586  ACTACACCATACGACGACAAATGCCATGGAACTCAGTACCAAGTATTGCTGCAGGTACA 645
Qy      80  GlySerThrAsn-----LysGlyMetAlaProGlnAlaAsnLeuValPheGlnSer 96
Db      646  GGTGCTGGAACACAGCTTTTACAAAGCGCTGCTCTGATGCTTTGTTGGTAGGAATAAAA 705
Qy      97  IleMetAspSerGlyGlyGly-----LeuGlyGlyLeuProSerAsn 110
Db      706  GTTTTAGTCAATGGAAGCGCAGCATGAGCAGTAACTGACAGAAATTGACTGGGCT 765
Qy      111  LeuGln-----ThrLeuPheSerGlnAla 118
Db      766  GTTCAAAATAAAGATGATACGGAATCAAAGTTATAAATTAAGCCTCGCCTCTCTACA 825
Qy      119  TyrSerAlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAsnGlyAlaTyr 138
Db      826  AGTTCTCATGGA-----ACTGACTTACTCTATTAGCAGTGAAT----- 864
Qy      139  ThrThrAspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPhe 158
Db      865  -----AGAGCAGTAGAT-----AGCGGTATTGTAGTAGTTGTA 897
Qy      159  AlaAlaGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsn 178
Db      898  GCACAGAGAAACTCTGGCCCTGCCAAATACACCATAGGGTCCCTGCTGCTGGGAAAAA 957
Qy      179  AlaIleThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsn 198
Db      958  GCCATAACAGTCGACGACCAATGGCAGATGTA-----GGTGAACCTGGCTTTAAC 1005

```

```

QY 199 ILeAsnHisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysPro 218
DB 1006 CTT-----GCAAGCTTTTCCAGCGCGGCTCTACTCTCTGACGGAAGATAAAACCT 1056
QY 219 AspValMetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSer 238
DB 1057 GACATTGGCGCCCGAGATATAATAACTGCGCAAG----- 1095
QY 239 SerPheTrpAlaAsnHisAspSerLysTyAlaTyMetGlyThrSerMetAlaThr 258
DB 1096 -----GCGAATTCGTAATGGATATGTAACATACAGCGGTACAAGCATGGCAACA 1146
QY 259 ProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIle 278
DB 1147 CTTTGTAGCAGAACTGTGCTCTTATGCTTAAC-----GCTATTCACAAATC 1197
QY 279 ThrProLysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspIleGlyLeu 298
DB 1198 ACTCCAAATGATGCA-----AAAATATAATAATGCTACTGCAAAAGCTGGGGCCT 1251
QY 299 GlyTyProAsnGlyAsnGlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsnVal 318
DB 1252 CCAAGCAAAACGTTGACTATGGTGCAGGAAGATTGGATGGCTATGAAGCTATAAGGTA 1311
QY 319 Ala----- 324
DB 1312 GCAGTAATTTAGAGAAATATAATTGATGTACCAAAATCATTTATATT-----TCA 1365
QY 325 SerSerLeuSerThrSerGlnLysAla---ThrTyrSerPheThrAlaThr---AlaGly 342
DB 1366 GGTATCTGCCCGGTTCTCGTACGATGATCTGGACTTTTAAACGCAAAATACAAAGT 1425
QY 343 LysProLeuLysIleSerLeuValTrpSerAsp 353
DB 1426 TATCCAAATCGCAATAACATTAATAATTCCTGAC 1458

RESULT 14
AAT61454
ID AAT61454 standard; DNA; 2539 BP.
XX AC AAT61454;
XX DT 06-OCT-1997 (first entry)
XX DE Streptomyces viridosporus dhpa gene.
XX KW asymmetric hydrolase; dhpa; 4-substituted-1,4-dihydropyridine;
XX KW derivative; Streptomyces viridosporus; ester; chiral; synthesis;
XX KW cardiovascular; treatment; hypertension; ischaemic heart disease; ds.
XX OS Streptomyces viridosporus.
XX FH Key Location/Qualifiers
XX FT CDS 338..2539
XX FT /tag= a
XX FT /note= "no stop codon given"
XX FT misc_feature 950..2509
XX FT /tag= b
XX FT /note= "encodes AAW13666"
XX PN WO9705243-A1.
XX PD 13-FEB-1997.
XX PF 30-JUL-1996; 96WO-JP002147.
XX PR 31-JUL-1995; 95JP-00212975.
XX PR 29-FEB-1996; 96JP-00067478.
XX PA (SAOC ) MERCIAN CORP.
XX AR Arisawa A, Matsufuji M, Tsuruta T, Dobashi K, Nakashima T;

```

```

PI Isshiki K, Yoshioka T;
XX WPI; 1997-145682/13.
DR P-PSDB; AAW13666, AAW13667.
XX PT Asymmetric hydrolase gene derived from Streptomyces viridosporus - acts
PT on 4-substituted-1,4-dihydropyridine derivatives to produce chiral
PT derivatives useful for synthesis of cardiovascular drugs.
XX Claim 3; Page 49-55; 78pp; Japanese.
XX This sequence is the Streptomyces viridosporus dhpa gene which encodes an
CC asymmetric hydrolase which acts on 4-substituted-1,4-dihydropyridine
CC derivatives. The enzyme allows the efficient conversion of 4-substituted-
CC 1,4-dihydropyridine esters to chiral partially hydrolysed derivatives,
CC for use in the synthesis of cardiovascular drugs suitable for the
CC treatment of e.g. hypertension and ischaemic heart disease
XX Sequence 2539 BP; 433 A; 959 C; 867 G; 280 T; 0 U; 0 Other;
SQ
Alignment Scores:
Pred. No.: 2,85e-17 Length: 2539
Score: 348.00 Matches: 129
Percent Similarity: 42.00% Conservative: 52
Best Local Similarity: 29.93% Mismatches: 167
Query Match: 15.49% Indels: 84
DB: 2 Gaps: 13
US-09-985-689A-1-COPY (1-434) x AAT61454 (1-2539)
QY 2 AspValAlaArgGlyIleValLysAlaAspValAlaGlnSerSerTyrGlyLeuTyrGly 21
DB 953 GACACGTCGTCGGCAGATCGCGCCCAAGGCTGTGTCGCGC---GGTCACACGCGC 1009
QY 22 GlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSer 41
DB 1010 AAGGCGGTGAAGATCGCGCTCTCGACACCGGTGTGACACGAGC----- 1054
QY 42 MetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsnAsn 61
DB 1055 ---CATCCGGACCTGAAGGCGCGGTGACCGCGTCCAAAGAACTTCACCGCGCGCGCGC 1111
QY 62 AlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGlySer 81
DB 1112 GCGCGGCAAGGTGGGCCACGGCACCCACGTCGCTCGATCGCGGGGCGGACGCGCGC 1171
QY 82 ThrAsn-----LysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIle 97
DB 1172 CAGTCCAAAGGCAAGTACAAGGCGCTCGCACCGCGCGCGCATCTCTCAACGCGCAAGTC 1231
QY 98 MetAspSerGlyGlyLeuGlyLeuGlyLeuProSerAsnLeuGlnThrLeuPheSerGln 117
DB 1232 CTCGACGACTCCGHT-----TTGCGGACGACTCCGCGCATCTCCGCGGATGAGTGG 1285
QY 118 AlaTySerAlaGlyAlaArgIle-HisThrAsnSerTrpGlyAlaAlaValAsnGlyAl 137
DB 1286 GCGGCGCGCAGGCGCGCGAGCTCGTCAATGACCTGGCGGCGATGACACACCGGAG 1345
QY 137 aTyThrThrAspSerArgAsnValAspTyrValArgLysAsnAspMetThrIleLe 157
DB 1346 ACCGACCCGCTGGAGGCGCGG-GTCGCAAGCTGCCCGCGAGAGGCGCTCTGCTTTCGC 1404
QY 157 uPheAlaAlaGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLy 177
DB 1405 CATCGCGCGCGCAACGAGGCGCGGAG-----TCGATCGTTCGCGCGCGCGCGGA 1458
QY 177 sAsnAlaIleThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAs 197
DB 1459 CGCCGCGCTCACCGTCGGCGCC-----GTGCA 1485
QY 197 pAsnIleAsnHisValAlaGlnPheSerSerArgGlyPro---ThrLysAspGlyArgIle 216
DB 1486 CGACAAAGGACAGCTCGCGGACTTCTCTCCACCGGCGCGCGCTCGCGGCGCGCAT 1545

```



```
Db 1286 GGGCGCGCGCAGGGCGCGCGCGTGTCAACATGAGCTGGCGGCGATGGACACACCGGAG 1345
QY 137 aTyrThrThrAspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLe 157
Db 1346 ACCGACCCCGTGGAGCGCGG-GTCGACAAAGTGTCCGCGAGAGAGGGCGTCTTTCGC 1404
QY 157 uPheAlaAlaGlyAsnGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLys 177
Db 1405 CATCGCGGCGCGCAACAGAGCGCGCGAG-----TCGATCGGTTCGCCCGGCGAGCGCGA 1458
QY 177 sAsnAlaIleThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAs 197
Db 1459 CGCGCGCTCACCGTCGGCGCC-----GTGGA 1485
QY 197 pAsnIleAsnHisValAlaGlnPheSerSerArgGlyPro-----ThrLysAspGlyArgIl 216
Db 1486 CGACAAAGGACAAAGCTCCGCGACTTCTCTCCACCGCGCGCGCTCCGCGACGCGGCCAT 1545
QY 216 eLysProAspValMetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaPr 236
Db 1546 CAAGCGGAGCGTCACCGCTCCCGGGGTGGACATCACGGCGCGCTCGGGGAGGGGCAACGA 1605
QY 236 oAspSerSerPheTrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMe 256
Db 1606 CATCGGCCAGGAGTCCGTGAGGGACCGCGCGGTACATGACCATCTCCGCGACGTCGAT 1665
QY 256 tAlaThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnAr 276
Db 1666 GGGACCCCGCGACGTCCGGCGCGCGCGCTCTCTGAAGCAGCAG----- 1711
QY 276 gGlyIleThrProLysProSerLeuLeuLysAlaAlaLeu-----IleAlaGlyAl 293
Db 1712 -----CACCCGCGACTGGACCTCCGCGCACTGAAGGGCGGCTCACCGGCTC 1758
QY 293 aAlaAspIleGlyLeuGlyTyrProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAs 313
Db 1759 CACCAAG---GGCGCAAGTACACCCCGTTCGAGCAGGGTTCGGGCGGATCCAGGCCGA 1815
QY 313 pLysSerLeuAsnValAlaTyrValAsnGluSerSerLeuSer----- 328
Db 1816 CAAGGCGCTCCAGCACACCGTGATCGCCGACCCGGTCTCGTGAGCTTCGGGCTCCAGCA 1875
QY 329 -----ThrSerGlnLysAlaThrTyrSerPheThrAlaTh 340
Db 1876 GTGGCGCGCACACCGACGAGCGGTCACCAAGCAGCTGACCTACCGCAACCTCGGCAC 1935
QY 340 rAlaGlyLysProLeuLysIleSerLeuValTrpSerAsp----- 353
Db 1936 CCAGGACGTACCGCTGAGCTGACGTGACCGCCACCGACCCCAAGGGCAAGCGGCGCCC 1995
QY 354 -----AlaProAlaSerThrAlaSe 361
Db 1996 GCGGGGCTTCTTCACGCTGGCGCCACACCGGTGACCGTCCCGGGCGGCGGCGGCGCTC 2055
QY 361 rValThrLeuValAsnAspLeuAsnLeu-----Va 371
Db 2056 CGTCGACATGACCGCGCACACCGGCTCGGGGCGACGCTGGGACGCTGGGCGGCTACTCGGCGTA 2115
QY 371 lIleThrAlaProAsnGlyThrGlnTyrVal 381
Db 2116 CGTGGTCGCCACGGCGGCGGCGGCGAGCGGTC 2146
```

Search completed: March 15, 2004, 20:47:42
Job time : 425 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: March 15, 2004, 20:31:52 ; Search time 83 Seconds
(without alignments)
2901.791 Million cell updates/sec

Title: US-09-985-689a-1-COPY
Perfect score: 2247
Sequence: 1 NDVARGIVKADVAQSSYGLY.....EVQAVNPVPGQTSLAIW 434

Scoring table:
BLOSUM62
Xgapop 10.0, Xgapext 0.5
Ygapop 10.0, Ygapext 0.5
Fgapop 6.0, Fgapext 7.0
Delop 6.0, Delext 7.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO.spool/SWOPE985689/runat_10032004_112807_19395/app_query.fasta_1.583
-DB=Issued Patents NA -OPMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blom62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=SWOPE985689 @CGN_1_1_44 @runat_10032004_112807_19395 -NCPUS=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents NA:
1: /cgn2_6/ptodata/2/ina/5A COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PTUS COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|--------|-------------|--------|----|-------------------|
| 1 | 2242 | 99.8 | 1923 | 4 | US-09-509-814A-5 |
| 2 | 2237 | 99.6 | 1923 | 4 | US-09-509-814A-7 |
| 3 | 2178 | 96.9 | 1920 | 4 | US-09-509-814A-3 |
| 4 | 2120.5 | 94.4 | 3003 | 2 | US-08-873-479-41 |
| 5 | 447.5 | 19.9 | 1977 | 3 | US-08-894-818B-2 |
| 6 | 447.5 | 19.9 | 1977 | 4 | US-09-445-472-11 |
| 7 | 411.5 | 18.3 | 1236 | 4 | US-09-445-472-2 |
| 8 | 411.5 | 18.3 | 1566 | 3 | US-08-894-818B-4 |
| 9 | 411.5 | 18.3 | 1962 | 3 | US-08-894-818B-34 |
| 10 | 411.5 | 18.3 | 1962 | 4 | US-09-445-472-15 |
| 11 | 398.5 | 17.7 | 1977 | 3 | US-08-894-818B-6 |
| 12 | 353 | 15.7 | 2539 | 3 | US-09-000-016-3 |

| | | | | | | |
|----|-------|------|-------|---|--------------------|--------------------|
| 13 | 353 | 15.7 | 2539 | 4 | US-09-514-340-3 | Sequence 3, Appli |
| 14 | 348 | 15.5 | 2809 | 3 | US-09-000-016-1 | Sequence 1, Appli |
| 15 | 348 | 15.5 | 2809 | 4 | US-09-514-340-1 | Sequence 1, Appli |
| 16 | 311.5 | 13.9 | 2532 | 1 | US-07-671-376C-4 | Sequence 4, Appli |
| 17 | 307 | 13.7 | 2835 | 1 | US-08-750-532-2 | Sequence 2, Appli |
| 18 | 307 | 13.7 | 4765 | 1 | US-08-750-532-8 | Sequence 8, Appli |
| 19 | 307 | 13.7 | 4765 | 3 | US-08-894-818B-7 | Sequence 7, Appli |
| 20 | 307 | 13.7 | 4765 | 4 | US-09-445-472-5 | Sequence 5, Appli |
| 21 | 302.5 | 13.5 | 1859 | 3 | US-08-894-818B-15 | Sequence 15, Appli |
| 22 | 296 | 13.2 | 898 | 1 | US-08-750-532-7 | Sequence 7, Appli |
| 23 | 290.5 | 12.9 | 564 | 1 | US-08-750-532-11 | Sequence 11, Appli |
| 24 | 290.5 | 12.9 | 564 | 3 | US-08-894-818B-14 | Sequence 14, Appli |
| 25 | 276 | 12.3 | 1306 | 4 | US-09-966-921A-1 | Sequence 1, Appli |
| 26 | 276 | 12.3 | 1330 | 4 | US-09-966-921A-5 | Sequence 5, Appli |
| 27 | 256.5 | 11.4 | 1194 | 4 | US-09-328-352-3407 | Sequence 3407, Ap |
| 28 | 256.5 | 11.4 | 10216 | 2 | US-08-875-154-1 | Sequence 1, Appli |
| 29 | 252 | 11.2 | 1628 | 4 | US-09-634-238-212 | Sequence 212, App |
| 30 | 251 | 11.2 | 840 | 1 | US-08-434-255-7 | Sequence 7, Appli |
| 31 | 251 | 11.2 | 840 | 1 | US-08-459-967-7 | Sequence 7, Appli |
| 32 | 251 | 11.2 | 840 | 1 | US-08-460-327-7 | Sequence 7, Appli |
| 33 | 251 | 11.2 | 840 | 1 | US-08-459-871-7 | Sequence 7, Appli |
| 34 | 251 | 11.2 | 840 | 3 | US-09-024-532-1 | Sequence 1, Appli |
| 35 | 251 | 11.2 | 840 | 4 | US-09-104-623A-1 | Sequence 1, Appli |
| 36 | 251 | 11.2 | 840 | 4 | US-09-019-532-1 | Sequence 1, Appli |
| 37 | 251 | 11.2 | 840 | 4 | US-09-417-359A-1 | Sequence 1, Appli |
| 38 | 251 | 11.2 | 840 | 4 | US-09-705-185-1 | Sequence 1, Appli |
| 39 | 251 | 11.2 | 840 | 4 | US-09-338-746-1 | Sequence 1, Appli |
| 40 | 251 | 11.2 | 1110 | 1 | US-08-434-255-5 | Sequence 5, Appli |
| 41 | 251 | 11.2 | 1110 | 1 | US-08-459-967-5 | Sequence 5, Appli |
| 42 | 251 | 11.2 | 1110 | 1 | US-08-460-327-5 | Sequence 5, Appli |
| 43 | 251 | 11.2 | 1110 | 1 | US-08-459-871-5 | Sequence 5, Appli |
| 44 | 251 | 11.2 | 1191 | 1 | US-08-434-255-3 | Sequence 3, Appli |
| 45 | 251 | 11.2 | 1191 | 1 | US-08-459-967-3 | Sequence 3, Appli |

ALIGNMENTS

RESULT 1
US-09-509-814A-5
; Sequence 5, Application US/09509814A
; Patent No. 6376227
; GENERAL INFORMATION:
; APPLICANT: TAKAIWA, MIKIO
; APPLICANT: OKUDA, MITSUYOSHI
; APPLICANT: SAKI, KATSUHIKA
; APPLICANT: KUBOTA, HIROMI
; APPLICANT: HITOMI, JUN
; APPLICANT: KAGEYAMA, YASUSHI
; APPLICANT: SHIKATA, SHITSUW
; APPLICANT: NOMURA, MASAFUMI
; TITLE OF INVENTION: ALKALINE PROTEASE
; FILE REFERENCE: 0327-0832-08CT
; CURRENT APPLICATION NUMBER: US/09/509,814A
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: PCT/JP98/04528
; PRIOR FILING DATE: 1998-10-07
; PRIOR APPLICATION NUMBER: JP 9-274570
; PRIOR FILING DATE: 1997-06-08
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 5
; LENGTH: 1923
; TYPE: DNA
; ORGANISM: Bacillus sp.
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1923)
US-09-509-814A-5

Alignment Scores:
Pred. No.: 2,92e-224 Length: 1923
Score: 2242.00 Matches: 433
Percent Similarity: 100.00% Conservative: 1

739 TCATGTCATGAAGCCTTCCGCGGGGAAAAATTACTGCATTATATGTCATTTGCGGACGGACGAAT 798
61 AsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
799 AATGCCAATGATACGAATGGTATGCTACGATGGCTGGCTCCGTTATTAGCAAAACGCG 858
81 SerThrAsnGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100
859 TCACCTAATAAAGGAATGGCGCTCAGCGGAATCTAGTCTTCCAACTCATCATGATGATGC 918
101 GlyGlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTySer 120
919 GGTGGGGGACTTGGAGGACTTACCTTCGAATCTGCAAACTTATTACGCAAGCATACAGT 978
121 AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAsnGlyAlaTyThrThr 140
979 GCTGGTGCCAGAATTATCAAACTCTCTGGGGAGCAGCAGTGAATGGGGCTTACACAACA 1038
141 AspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAla 160
1039 CATTCAGAAATGTGGATGCTATGTGGCAAAATGATATGACGATCCTTTTCGTGCGC 1098
161 GlyAsnGlyProAsnGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 180
1099 GGGAAATCAGAGACCGAAGCGGAACCATCAGTGCACACGACGACGCTAAAAATGCAATA 1158
181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 200
1159 ACAGTCGAGCTACGGAAACCTCCGCCCAAGCTTTGGGTCTTATCGGACCATATCAAC 1218
201 HisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVal 220
1219 CATGTGGCACAGTCTCTTCCACGTGGACCGACAAAGGATGGACGGATCAAAACGGATGTC 1278
221 MetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerPhe 240
1279 ATGGCACCGGAGCTTACATACATATCAGCAAGATCTTCTTGCACCGGATCCTCCTTC 1338
241 TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyThrSerMetAlaThrProIle 260
1339 TGGCGCAACCATGACATGAATAATGATCATATCGGTGGACGCTCCATCGCTACACCGATC 1398
261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPro 280
1399 GTTGTCTGGAACGTGGCACAGCTCTGTGAGCATTTTGTGAAAAACAGAGCATCACACCA 1458
281 LysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspIleGlyLeuGlyTyr 300
1459 AAGCTTCTCTATTAAAGCGGCACATGATTCGCGGTGCACCTGACATCGGCTTGGCTAC 1518
301 ProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaTy 320
1519 CCGAACGGTAACCAAGGATGGGACGAGTGCATTTGGATAAATCCCTGAACGTTGCTAT 1578
321 ValAsnGluSerSerSerLeuSerThrSerGlnLysAlaThrTyrSerPheThrAlaThr 340
1579 GTGAACAGTCCAGTCTCTATCCACCGCCAAAGACGACGCTACTCTGTTTACTGCTACT 1638
341 AlaGlyLysProLeuLysIleSerLeuValTrpSerAspAlaProAlaSerThrThrAla 360
1639 GCCGGCAAGCTTTGAAAACTCCCTGGTATGTTCTGATGCCCTCGAGCACAACTGCT 1698
361 SerValThrLeuValAsnAspLeuAsnLeuValIleThrAlaProAsnGlyThrGlnTyr 380
1699 TCCGTAAACGCTTGCAATGATCTGGACCTTGTATATACCCCTCAATATGGCACACAGTAT 1758
381 ValGlyAsnAspPheThrSerProTyrAsnAspAsnTrpAspGlyArgAsnAsnValGlu 400
1759 GTAGGAATGACTTACTTCGCCATACATGATAACTGGGATGGCGCAATAACGTAGAA 1818
401 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyAsn 420
1819 AATGTATTATTAAATGACCACCAAGCGGACGCTATACAAATTGAAGTACAGGCTTATAAC 1878

Qy 421 ValProValGlyProGlnThrPheSerLeuAlaIleValAsn 434
Db 1879 GTACCGTGGACCACAGAACTTCTCGTTGGCAATTGTGAAT 1920

RESULT 3

US-09-509-814A-3
; Sequence 3, Application US/09509814A
; Patent No. 6376227
; GENERAL INFORMATION:
; APPLICANT: TAKAIWA, MIKIO
; APPLICANT: OKUDA, MITSUYOSHI
; APPLICANT: SAEKI, KATSUHIKA
; APPLICANT: KUBOTA, HIROMI
; APPLICANT: HITOMI, JUN
; APPLICANT: KASEYAMA, YASUSHI
; APPLICANT: SHIKATA, SHITSUW
; APPLICANT: NOMURA, MASAFUMI
; TITLE OF INVENTION: ALKALINE PROTEASE
; FILE REFERENCE: 0327-0832-0PCT
; CURRENT APPLICATION NUMBER: US/09/509,814A
; CURRENT FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: PCT/JP98/04528
; PRIOR FILING DATE: 1998-10-07
; PRIOR APPLICATION NUMBER: JP 9-274570
; PRIOR FILING DATE: 1997-06-08
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 3
; LENGTH: 1920
; TYPE: DNA
; ORGANISM: Bacillus sp.
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1920)
US-09-509-814A-3

Alignment Scores:

Pred. No.: 1,41e-217 Length: 1920
Score: 2178.00 Matches: 417
Percent Similarity: 99.31% Conservative: 14
Best Local Similarity: 96.08% Mismatches: 3
Query Match: 96.93% Indels: 0
DB: 4 Gaps: 0

US-09-985-689A-1-COPY (1-434) x US-09-509-814A-3 (1-1920)

Qy 1 AsnAspValAlaAlaArgGlyIleValLysAlaAspValAlaGlnSerSerTyrGlyLeuTyr 20
Db 616 AATGATGTGGCCACAGAGGTATTGTCAAAGCGGATGTGGCACAGACGACGTACGGTTTGTAT 675
Qy 21 GlyGlnGlyGlnIleValAlaAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db 676 GGACAAGCCAGATGTGCGCAGTTGCCGATACTCGATTTGGATACAGGAAGAACGACAGT 735
Qy 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
Db 736 TCGATCATGAGCGCTTCGCGGTAAATAACACACCATATATGCTGGGTCCGACGAAT 795
Qy 61 AsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
Db 796 AATGCGAATGATACGAACCGTCATGCCATCCCATCGGATTCGGTATTAGGAAATGGC 855
Qy 81 SerThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100
Db 856 GCAACCAATTAAGGATGGACCTCAAGCGAATCTGTTTTTCAATCCATCATGATGATGC 915
Qy 101 GlyGlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTySer 120
Db 916 AGTGTGGGCTTGGAGGCTTGCCTTCCAATCTGCAAACTTATTTCAGCAAGCATTCAGT 975
Qy 121 AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAsnGlyAlaTyThrThr 140

976 GCAGGTGCCAGAAATTCATACAAATCTCTGGGGGCGAGCGTGAATGGGGCTCACACGACA 1035
Qy 141 AspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAala 160
Db 1036 GATTCAGAAATGCGATGACTATGTAGGAAAATGATATGACGATCTCTTTGCGGCT 1095
Qy 161 GlyAsnGluGlyProAsnGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 180
Db 1096 GGGAAATGAAGCGGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1155
Qy 181 ThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 200
Db 1156 ACAGTCGGCGCAACCGCAAACTGCTCGACCTTCGGTTCATATGCGAATATTAAC 1215
Qy 201 HisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVal 220
Db 1216 CAGCTTGACAGCTCTCTCCCGTGGCGCGACAAAAGATGGCGGAATCAAGCTGATGTC 1275
Qy 221 MetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPhe 240
Db 1276 ATGGCGCCAGGACATACATATTTATCAGCAAGATCTCTTTCACCCGATTCCTCTTC 1335
Qy 241 TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyThrSerMetAlaThrProIle 260
Db 1336 TGGCGGAATCATGACAGCAATATGCTATATGGTGGAACTGCTCATGCGCAACCGGAT 1395
Qy 261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPro 280
Db 1396 GTTGGGGGAATGTTGCACAGCTCGGTGAGCAATTTTGTGAAAATAGAGGAATCACTCT 1455
Qy 281 LysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspIleGlyLeuGlyTyr 300
Db 1456 AAGCCTTCCTATTGAAGCAGCTTGTATGTCAGTGTGCTGTGATGTTGGATGGTTAT 1515
Qy 301 ProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 320
Db 1516 CCGAACCGAAACCAAGATGGGGCGAGTGCACCTCGGATGATGATGATGATGATGAT 1575
Qy 321 ValAsnGluSerSerSerLeuSerThrSerGlnLysAlaAlaThrTyrSerPheThrAlaThr 340
Db 1576 GTGAACGAATCCAGTCCCTATCACTAGCCAAAAGCGACATATACCTTACTGCAACG 1635
Qy 341 AlaGlyLysProLeuLysLysSerLeuValTrpSerAspAlaProAlaSerThrAla 360
Db 1636 GCGGGCAAGCAATGAAATCTCCCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1695
Qy 361 SerValThrLeuValAsnAspLeuAsnLeuValIleThrAlaProAsnGlyThrGlnTyr 380
Db 1696 TCTGTAACTCGTCAATGATTTGGATTTGGTTCATTTACAGCACCAACCGAACAGATAT 1755
Qy 381 ValGlyAsnAspPheThrSerProTyrAsnAspAsnTrpAspGlyArgAsnValGlu 400
Db 1756 GTCGGGAATGACTTCTCAGCACCATTTGACAAATACTGGGATGGCGCAATAACGTAGAA 1815
Qy 401 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 420
Db 1816 AATGTTATTTATTAATTCGCCCAACCAAGTGAACATATACCATTTGAGTGTGACATATAT 1875
Qy 421 ValProValGlyProGlnThrPheSerLeuAlaIleValAsn 434
Db 1876 GTCCCGGTTGACACCAAAATCTCTCGTTGGCAATTTGTGAAC 1917

RESULT 4

US-08-873-479-41
; Sequence 41, Application US/08873479
; Patent No. 5891701
; GENERAL INFORMATION:
; APPLICANT: Sloma, Alan
; APPLICANT: Lytne, Christianson
; TITLE OF INVENTION: Nucleic Acids Encoding A Polypeptide
; TITLE OF INVENTION: Having Protease Activity
; NUMBER OF SEQUENCES: 57
; CORRESPONDENCE ADDRESS:

ADDRESSEE: No. 5891701o No. 5891701disk of No. 5891701th America
STREET: 405 Lexington Avenue
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10174
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/873,479
FILING DATE: 12-JUN-1997
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Agtis, Cheryl H
REGISTRATION NUMBER: 34,086
REFERENCE/DOCKET NUMBER: 5251.000-US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-867-0123
TELEFAX: 212-878-9655
TELEX:
INFORMATION FOR SEQ ID NO: 41:
SEQUENCE CHARACTERISTICS:
LENGTH: 3003 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
US-08-873-479-41

Alignment Scores:
Pred. No.: 2 91e-211 Length: 3003
Score: 2120.50 Matches: 405
Percent Similarity: 97.93% Conservative: 20
Best Local Similarity: 93.32% Mismatches: 8
Query Match: 94.37% Indels: 1
DB: 2 Gaps: 1

US-09-985-689A-1-COPY (1-434) x US-08-873-479-41 (1-3003)
Qy 1 AsnAspValAlaArgGlyIleValLysAlaAspValAlaGlnSerSerTyrGlyLeuTyr 20
Db 1470 AATGACGTGGCCGTGGCATTTGTGAAGCAGACGTGCGCAAAATACCTTTGCTTATAT 1529
Qy 21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db 1530 GGCAAGGACAGATTTGATGAGTGTGATGATGATGATGATGATGATGATGATGAT 1589
Qy 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsn 60
Db 1590 TCGATGATGAAGCATTTCCCGGTAAAGATTACCGCACTATATGCTGGCGCAGACGAT 1649
Qy 61 AsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
Db 1650 AACCCCAATGATCCCAATGACATGGACCCCATGTTGCTGGATCTGTGTAGGAAT--- 1706
Qy 81 SerThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100
Db 1707 GCTACAAATAAAGGATGGCAGCCGCAAGCCCAATCTAGTCTTTCAATCTATTATGGAT 1766
Qy 101 GlyGlyGlyLeuGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyrSer 120
Db 1767 GGTGGAGGCTGGAGGACTTCTGCTATCTACAAACATTTATTCAGTCAAGCATATAGT 1826
Qy 121 AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAsnGlyAlaTyrThr 140
Db 1827 GCTGAGCGAATTCATACGAATTCATGGGGGCTCCAGTAAACCGTGCCTATACGACA 1886
Qy 141 AspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAla 160
Db 1887 GACTCTCGAAATGTTGATGATTATGTGAGAAAAATGATATGACGATCTTTTTTGGGCC 1946

QY 161 GlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAla 180
Db 1947 GGAATGAGGACGAGTAGCGTACAAATCAGTGCACGAGCAACAAAATGCGATT 2006
QY 181 ThrValGlyValaThrGluAsnLeuArgProSerPheGlySerTyraAlaAspAsnIleAen 200
Db 2007 ACAGTTGGGCAACCGAACCACCTACGTCACAGTTCCGATCTATCGGATATATTAC 2066
QY 201 HisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVal 220
Db 2067 CATGTTGCTCAATCTCTTCCAGAGTCTCTACTAGAGATGGAGTATTAAAGCGGACGTC 2126
QY 221 MetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerPhe 240
Db 2127 ATGGCACCGAGTACGATATCTCTCTGCTAGATCATCATCTAGCTCCAGATCTCTCATTC 2186
QY 241 TrpAlaAsnHisAspSerTyraAlaTyraMetGlyGlyThrSerMetAlaThrProIle 260
Db 2187 TGGCAAAACCATGATGATGAATATGCTACATCGGTGGTACTTCTTATGCTACTCCAAAT 2246
QY 261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgIleThrPro 280
Db 2247 GTAGCAGGTATGTTGCACAAATTAAGGGAGCATTTTGTGAAAATAGAGGGGTAACTCT 2306
QY 281 LysProSerLeuLysAlaAlaLeuIleAlaGlyValaAlaAspIleGlyLeuGlyTyr 300
Db 2307 AAGCTTCCTTTTAAAGCTGCTTTAAATGCGAGTGTCTGGATCTTGGACTTGGCTTT 2366
QY 301 ProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 320
Db 2367 CCAATGTTAACCAAGGATGGGAAGAGTAACGTTAGATAAATCCCTAAATGTGCGAAT 2426
QY 321 ValAsnGluSerSerSerLeuSerThrSerGlnLysAlaThrTyrSerPheThrAlaThr 340
Db 2427 GTGATGAAAGAGCCCTTTATCAACAGTCAAAAGCAACATTTCTGTTTACGGTCAA 2486
QY 341 AlaGlyLysProLeuLysIleSerLeuValTrpSerAspAlaProAlaSerThrAla 360
Db 2487 GCTGGTAAACCTTTAAATATATACATTTGTTGGTCAAGTCCACAGTAGCAGCAGCGCA 2546
QY 361 SerValThrLeuValAsnAspLeuAsnLeuValIleThrAlaProAsnGlyThrGlnTyr 380
Db 2547 TCACATACTTTAGTGAATGATTTAGACTTAGTATGATCTACCAATGGAATGAATATAC 2606
QY 381 ValGlyAsnAspPheThrSerProTyraAsnAspAsnTrpAspGlyArgAsnValGlu 400
Db 2607 GTCGAAATGACTTTACAGCACCGTATGATACAAATTTGGGATGGCAGAAACACGTTGAA 2666
QY 401 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 420
Db 2667 AATGTGTTTATCAATGCTCTCAAGCGGAACGTTATACATGCAAGTGCAGGCTTACAT 2726
QY 421 ValProValGlyProGlnThrPheSerLeuAlaIleValAsn 434
Db 2727 GTACCAAGTACGCGCAACCTTTTCTTTAGCGATTGTACAT 2768

RESULT 5

US-08-894-818B-2

; Sequence 2, Application US/08894818B

; Patent No. 6261822

; GENERAL INFORMATION:

; APPLICANT: TAKAKURA, Hikaru

; APPLICANT: MORISHITA, Mio

; APPLICANT: YAMAMOTO, Katsuhiko

; APPLICANT: MITTA, Masanori

; APPLICANT: ASADA, Kiyozo

; APPLICANT: TSUNASAWA, Susumu

; APPLICANT: KATO, Ikunoshin

; TITLE OF INVENTION: HYPER-THERMOSTABLE PROTEASE GENES

; NUMBER OF SEQUENCES: 42

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Browdy and Neimark

STREET: 419 Seventh Street N.W., Ste. 300
CITY: Washington
STATE: D.C.
COUNTRY: United States of America
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/894,818B
FILING DATE: 20-MAY-1998
CLASSIFICATION: 435
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: PCT/JP96/03253
FILING DATE: 07-NOV-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 323285/1995
FILING DATE: 12-DEC-1995
ATTORNEY/AGENT INFORMATION:
NAME: Browdy, Roger L.
REGISTRATION NUMBER: 25,618
REFERENCE/DOCKET NUMBER: TAKAKURA-1
TELEPHONE: (202) 628-5197
TELEFAX: (202) 737-3528
INFORMATION FOR SEQ ID NO: 2:
LENGTH: 1977 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: genomic DNA
US-08-894-818B-2

Alignment Scores:

Pred. No.: 8,24e-37 Length: 1977

Score: 447.50 Matches: 137

Percent Similarity: 44.66% Conservative: 68

Best Local Similarity: 29.85% Mismatches: 153

Query Match: 19.92% Indels: 101

DB: 3 Gaps: 18

US-09-985-689A-1-COPY (1-434) x US-08-894-818B-2 (1-1977)

QY 8 ValLysAlaAspValAlaGlnSerSerTyrglyLeuTyrglyGlnIleValAla 27
Db 433 ATAGGGGCGGATACCGTCTGGAACTCCCTCGGTACGACGAGCGGTGTGGTGGTCC 492
QY 28 ValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisGluAlaPheArg 47
Db 493 ATCGTCGATACGGGTATAGCGGAAC-----CACCCGATCTGAG 534
QY 48 GlyLysIleThrAlaLeuTyr---AlaLeuGlyArgThrAsnAsnAlaAsnAspThrAsn 66
Db 535 GCGAAGGTCATAGGTGCTGACGCGCGTCAACGCGAGTCCGCCCTACGATGACCAG 594
QY 67 GlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGlySerThrAsnLys----- 84
Db 595 GGACACGGAACCCACGTTGCGGGTATCGTCCCGGAACCGGACGCGTTAACTCCAGTAC 654
QY 85 ---GlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMet-----AspSerGly 101
Db 655 ATAGCGGTGCGCGCGCGCGGAGTCTCGCGGTCAAGGTCTCGGTGCGGACGGTTCG 714
QY 102 GlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyrSerAla 121
Db 715 GGAAGCGTCTCCACCATCATCGCGGTGTTGACTGGGTCTCCAGAACCAAGCAAGTAC 774
QY 122 GlyAlaArgIle-----HisThrAsnSer 129
Db 775 GGGATAAGGTCATCAACTCTCCCTCGGCTCTCCAGAGCTCCGACGGAACCGACTCC 834

QY 130 TrpGlyAlaAlaValAsnGlyAlaTyrThrThrAspSerArgAsnValAspAspTyrVal 149
DB 835 CTCAGTCAGGCGCTCAACAGCGCTGGACGCC----- 867
QY 150 ArgLysAsnAspMetThrLeuPheAlaAlaGlyAsnGluGlyProAsnGlyGlyThr 169
DB 868 -----GGTATAGTAGTCTGGCTGCCCGCGGCAACAGCGCGGCAACCTACACC 918
QY 170 IleSerAlaProGlyThrAlaLysAsnAlaIleThrValGlyAlaThrGluAsnLeuArg 189
DB 919 GTCGGCTCACCGCGCGGCGAGGAGGTCTAATACCGTGGTGCA----- 963
QY 190 ProSerPheGlySerTyrAlaAspAsnIleAsnHisValAlaGlnPheSerSerArgGly 209
DB 964 -----GTTGACAGCAACAGCAACATCGCCAGCTTCTCCAGCAGGGGA 1005
QY 210 ProThrLysAspGlyArgLysProAspValMetAlaProGlyThrPheIleLeuSer 229
DB 1006 CGGACCGCGGACGGAAGGCTCAAGCGGAAGTCTGGCGCGGCGGTGGACATCATAGCC 1055
QY 230 AlaArgSerSerLeuAlaProAspSerPheTrpAlaAsnHisAspSerLysTyrAla 249
DB 1066 CCGCGCGCGCAGC-----GGAACCGAGCATGGCCACCCCGATAAACGACTACTACACC 1116
QY 250 TyrMetGlyGlyThrSerMetAlaThrProIleValAlaGly-----AsnValAlaGlnLeu 268
DB 1117 AAGGCTCTGGAACAGCATGGCCACCCCGCAGCTTTCGGCGGTGGCGCTCATCCTC 1176
QY 269 ArgGluHisPheValLysAsnArgGlyLeuThrProLys-----ProSerLeuLeuLys 286
DB 1177 CAGGCGCCAC-----CCGAGCTGGACCCCGGCAAGGTGAAG 1212
QY 287 AlaAlaLeuIleAlaGlyAla-----AlaAspIleGlyLeu 298
DB 1213 ACGGCTCATCAGACCGCGGACATAGTGGCCCAAGAGATAGCGGACATCGCTAC 1272
QY 299 GlyTyrProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeu----- 316
DB 1273 GTGCGC-----GGTAGGGTGAACGTCTACAAGGCCATCAAGTAC 1311
QY 317 ---AsnValAlaTyrValAsnGluSerSerSerLeuSerThrSerGlnLysAlaThrTyr 335
DB 1312 GAGGACTAGCCCAAGCTCACCTTACCGGCTCCGTCGCGGCAAGGAGCGCCACCCAC 1371
QY 336 SerPheThrAlaThrAlaGlyLysProLeuLysIleSerLeuValTrpSerAspAlaPro 355
DB 1372 ACCTTCGACGTACGCGGCGCCACCTTCGTGACCGCCACCCCTCTACTGGAC----- 1422
QY 356 AlaSerThrThrAlaSerValThrLeuValAsnAspLeuAsnLeuValIleThrAlaPro 375
DB 1423 -----ACGGGCTGAGGACATCGACCTCTACCTCTACGACCC 1461
QY 376 AsnGlyThrGlnTyrValGlyAsnAspPheThrSerProTyrAsnAspTrpAspGly 395
DB 1462 AACGGGAACGAG-----GTTGACTACTCTCTACCGGCTACTAC----- 1500
QY 396 ArgAsnAsnValGluAsnValPheIleAsnAlaProGlnSerGlyThrThrIleGlu 415
DB 1501 -----GGCTTCGAGAGTGGCTACTACACCCGCGGCAACCTGGAGCGTCAAG 1554
QY 416 ValGlnAlaTyrAsnValProValGlyProGlnThrPheSerLeuAlaIleValAsn 434
DB 1555 GTCGTACGTACTAAG-----GGCGGGGGAAGTACACGAGTGGAGCTGCTGACG 1602

RESULT 6

US-09-445-472-11
; Sequence 11, Application US/09445472
; Patent No. 6358726
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: SHIMOJO, Tomoko

; APPLICANT: ASADA, Kiyozo
; APPLICANT: KATO, Ikuroshin
; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
; FILE REFERENCE: TAKAKURA-6
; CURRENT APPLICATION NUMBER: US/09/445,472
; CURRENT FILING DATE: 1999-12-06
; PRIOR APPLICATION NUMBER: 151969/1997
; PRIOR FILING DATE: 1997-06-10
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 11
; LENGTH: 1977
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
US-09-445-472-11

Alignment Scores: 8,24e-37 Length: 1977
Pred. No.: 447.50 Matches: 137
Score: 44.66% Conservative: 68
Percent Similarity: 29.85% Mismatches: 153
Best Local Similarity: 19.92% Indels: 101
Query Match: 4 Gaps: 18
DB:

US-09-985-689A-1-COPY (1-434) x US-09-445-472-11 (1-1977)

QY 8 ValLysAlaAspValAlaGlnSerSerTyrGlyLeuTyrGlyGlnGlyGlnIleValAla 27
DB 433 ATAGGGCGCGATACCGTCTGGAACCTCCCTCGGTACGACGGAAGCGGTGTGGTGTGCC 492
QY 28 ValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisGluAlaPheArg 47
DB 493 ATCGTCGATACGGGTATAGACGCGAAC-----CACCCCATCTGAAG 534
QY 48 GlyLysIleThrAlaLeuTyr---AlaLeuGlyArgThrAsnAsnAlaAsnAspThrAsn 66
DB 535 GGCAGGTCATAGGCTGGTACGACGCGCTCAACGCGGAGGTTCGACCCCTACGATGACCAG 594
QY 67 GlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGlySerThrAsnLys----- 84
DB 595 GGACACGGAACCCACGCTTCGGGTATCGTTCGCGGAACCGGCGAGGTTAACTCCAGTAC 654
QY 85 ---GlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMet-----AspSerGly 101
DB 655 ATAGGCGTCCCGCGCGGAGCTCGTGGCGTCAAGTTCGCGTCCGCGCGGTTCG 714
QY 102 GlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyrSerAla 121
DB 715 GGAAGCGTCTCCACCATCATCGCGGTGTGACTGGTCTGTCAGAACAGGACAAGTAC 774
QY 122 GlyAlaArgIle-----HisThrAsnSer 129
DB 775 GGATATAGGTCTATCACTCTCCCTCGGCTCTCCAGAGCTCCGACGGAACCGACTCC 834
QY 130 TrpGlyAlaAlaValAsnGlyAlaTyrThrThrAspSerArgAsnValAspAspTyrVal 149
DB 835 CTCAGTCAGGCGCTCAACACGCGCTGGACGCC----- 867
QY 150 ArgLysAsnAspMetThrIleLeuPheAlaAlaGlyAsnGluGlyProAsnGlyGlyThr 169
DB 868 -----GGTATAGTAGTCTGGCTGCCCGCGGCAACAGCGCGGCGGACCTACACC 918
QY 170 IleSerAlaProGlyThrAlaLysAsnAlaIleThrValGlyAlaThrGluAsnLeuArg 189
DB 919 GTCGGCTCACCGCGCGGCGGAGGTCATTAACCGTGGTGCA----- 963
QY 190 ProSerPheGlySerTyrAlaAspAsnIleAsnHisValAlaGlnPheSerSerArgGly 209
DB 964 -----GTTGACAGCAACAGCAACATCGCCAGCTTCTCCAGCAGGGGA 1005
QY 210 ProThrLysAspGlyArgLysProAspValMetAlaProGlyThrPheIleLeuSer 229

```
Db 1006 CCGACCGCGGAGGCTCAAGCGGAGTCTGCGCCCGCGGCTTGACATCATAGCC 1065
QY 230 AlaArgSerSerLeuAlaProAspSerPheTrpAlaAsnHisAspSerLysTrpAla 249
Db 1066 CCGCGCGCGCAGC-----GGACACGATGGGCACCCCGGATAAACGACTACTACACC 1116
QY 250 TyrMetGlyThrSerMetAlaThrProIleValAlaGly---AsnValAlaGlnLeu 268
Db 1117 AAGGCGCTCGGAACACGATGGCCCGCGACGTTTCGGCGGTGGCGCGCTCATCTC 1176
QY 269 ArgGluHisPheValLysAsnArgGlyIleThrProLys-----ProSerLeuLeuLys 286
Db 1177 CAGGCCAC-----CCGAGCTGGACCCCGGACAAAGTGAAG 1212
QY 287 AlaAlaLeuIleAlaGlyAla-----AlaAspIleGlyLeu 298
Db 1213 ACCGCGCTCATCGAGACCGCGACATAGTCGCCCCCAAGGAGATACGGGACATCGCCTAC 1272
QY 299 GlyTyrProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeu----- 316
Db 1273 GGTGCG-----GGTAGGTGAAGCTCTACAAAGCCATCAAGTAC 1311
QY 317 ---AsnValAlaTyrValAsnGluSerSerLeuSerThrSerGlnLysAlaThrTyr 335
Db 1312 GACGACTAAGCCAAAGCTCACTTCAACCGGCTCGCTCGCGACAAAGGAAGCGCCACCCAC 1371
QY 336 SerPheThrAlaThrAlaGlyLysProLeuLysIleSerLeuValTrpSerAspAlaPro 355
Db 1372 ACCTTCAGCTAGCGCGCCACCTTCGTGACCGCCACCTCTACTGGAC----- 1422
QY 356 AlaSerThrThrAlaSerValThrLeuValAsnAspLeuAsnLeuValIleThrAlaPro 375
Db 1423 -----ACGGGCTCGAGCGACATCGACCTCTACCTCTACGACCC 1461
QY 376 AsnGlyThrGlnTyrValGlyAsnAspPheThrSerProTyrAsnAspAsnTrpAspGly 395
Db 1462 AACGGGAACGAG---GTTGACTACTCTACACCGCCTACTAC----- 1500
QY 396 ArgAsnAsnValGluAsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGlu 415
Db 1501 -----GGCTTCGAGAGGTGCGCTACTACAACCCGACCGCGGAACCTGGACGGTCAAG 1554
QY 416 ValGlnAlaTyrAsnValProValGlyProGlnThrPheSerLeuAlaIleValAsn 434
Db 1555 GTCGTCAGCTACAAG-----GGCGCGGCGAAGTACACGAGTGCAGCTCGTCAGC 1602
```

RESULT 7

```
US-09-445-472-2
; Sequence 2, Application US/09445472
; Patent No. 6358726
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: SHIMOJO, Tomoko
; APPLICANT: ASADA, Kiyozo
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
; FILE REFERENCE: TAKAKURA-6
; CURRENT APPLICATION NUMBER: US/09/445,472
; CURRENT FILING DATE: 1999-12-06
; PRIOR APPLICATION NUMBER: 151969/1997
; PRIOR FILING DATE: 1997-06-10
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Patent in version 3.0
; SEQ ID NO 2
; LENGTH: 1236
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
US-09-445-472-2
```

```
Alignment Scores: 2,25e-33 Length: 1236
Pred. No.: 411.50 Matches: 140
Score: 42.55% Conservative: 60
Percent Similarity: 29.79% Mismatches: 148
Best Local Similarity: 18.31% Indels: 122
Query Match: 4 Gaps: 20
DB: 4

US-09-985-689A-1-COPY (1-434) x US-09-445-472-2 (1-1236)
QY 12 ValAlaGlnSerSerTyrGly-LeuTyr-----GlyGlnGlyG1 24
Db 24 GTCTGCAGCTCAAGTTATGCAACTTACGTTTGGAACTTGGGATATGATGGTCTCTGAAT 83
QY 24 nIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisG1 44
Db 84 CACATAGGATATGACACTGGAATTGAC-----GCTTCTCATCC 125
QY 44 uAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsnAsnAlaAsn-- 63
Db 126 AGATCTCCAAAGAAAGTA-----ATTGGTGGGTAGATTGTCATGG 170
QY 64 -----AspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuG1 78
Db 171 TAGAGATTATCCATACGATGACCATGGACATGGAACCTCATGCTTCAATAGCAGCTGG 230
QY 78 yAsnGlySerThrAsn-----LysGlyMetAlaProGlnAlaAsnLeuValPh 94
Db 231 TACTGGAGCAGCAAGTAATGCAAGTACAAGGAATGCTCCAGGAGCTAAGCTGGCGGG 290
QY 94 eGlnSerIleMet-----AspSerGlyGlyGlyLeuGlyLeuProSerAsnLeuG1 112
Db 291 AATTAAGGTTCTAGTCCGATGTTCTGGAAGCATATCTACTATAATTAAAGGAGTTGA 350
QY 112 nThrLeuPheSerGlnAlaTyrSerAlaGlyAlaArgIleHisThrAsnSerTrpGlyAl 132
Db 351 GTGGCGCGTTGATTAACAAGATAAGTACGGAATTAAAGTTCATTAACTTCTCTGTTTC 410
QY 132 a-----AlaValAsnGlyAlaTyrThrTh 140
Db 411 AAGCCAGAGCTCAGATGGTACTAGCTCTAAGTACAGCTGTTAATCCAGCGGGATGC 470
QY 140 rAspSerArgAsnValAlaAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAl 160
Db 471 T-----GGATTAGTCTGTTGTTGCTGCGC 494
QY 160 aGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaI1 180
Db 495 TCGAAACAGTGGACCTAACCAAGTATACATCGGTTCTCCAGCAGCTCAGCAAGATTAT 554
QY 180 eThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAs 200
Db 555 TACAGTTGGAGCC-----GTTGACAAGTATGA 581
QY 200 rHisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVa 220
Db 582 TGTATATAACAAGCTTCTCAAGCAGAGCGCCAACTGCAGCGCAGGCTTAAGCCTGAGGT 641
QY 220 lMetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPh 240
Db 642 TGTTCCTCCAGAACTGGATTAATGCTGCCAGCAGT-----GGAACATGAT 692
QY 240 eTrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProI1 260
Db 693 GGGTCAACCAATTAATGACTATTACACAGCAGCTCTCTGGGACATCAATGCAACTCTCA 752
QY 260 eValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPr 280
Db 753 CGTAGCTGGTATTGAGCCCTCTTGTCTCCAA-----GCACACCC 791
QY 280 oLys-----ProSerLeuLeuLysAlaAlaLeuIleAlaGlyAla----- 293
Db 792 GAGCTGGACTCCAGACAAAGTAAACACAGCCCTCATAGAAACTGCTGATATCGTAAAGCC 851
```


Db 642 TGTGTCTCCAGAACTGGATAATTCTGCCAGCAAGT-----GGAACTAGCAT 692
Qy 240 eNTPAlaAenHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrPro11 260
Db 693 GGGTCAACCAATTAATGATATATACACAGCACTCTGGAGATCAATGGCACTCTCA 752
Qy 260 eValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyLeuThrPr 280
Db 753 CGTAGCTGGTATTGACGCCCTCTTGCTCCAA-----GCACACCC 791
Qy 280 olys-----ProSerLeuLeuLysAlaAlaLeuLeuAlaGlyAla----- 293
Db 792 GAGCTGGATCCAGACAAAGTAAAGACAGCCCTCATGAACTGCTGATATCGTAAAGCC 851
Qy 294 -----AlaAspIleGlyLeuGlyTyrProAsnGlyAsnGlnGlyTyrGlyArgVa 310
Db 852 AGATGAATAGCCGATATAGCTTACGGTGCA-----GGTAGGGT 890
Qy 310 lThrLeuAspLysSerLeuAsnValAlaTyrValAsnGluSerSerSerLeuSerThrSe 330
Db 891 TAATGATACAGGCTATAAAC-----TACGATAACTATGCAAGCTAGTGTCTACCTGG 944
Qy 330 rGlnLysAla-----ThrTyrSerPheThrAlaThrAlaGlyLysProbe 345
Db 945 ATATGTGTGCCAACAAAGGCGAGCAAACTCACCAGTTCTGTTATTAGCGGAGCTTCGTGCT 1004
Qy 345 uLysIleSerLeuValTyrSerAspAlaProAlaSerThrThrAlaSerValThrLeuVa 365
Db 1005 AACTGGCCACATATTACTGGACAACTCCAAAT----- 1035
Qy 365 lAsnAspLeuAsnLeuValIleThrAlaProAsnGlyThrGlnTyrValGlyAsnAspPh 385
Db 1036 -AGCGACTTGAATCTTACCTTACGATCCCAATGGAACCCAG---GTTGACTACTCTTA 1091
Qy 385 eThrSerProTyrAsnAspAsnTrpAspGlyArgAsnAsnValGluAsnValPheIleAs 405
Db 1092 CACCGCTACTAT-----GGATTCGAAAAGGTTGGTTATTA 1127
Qy 405 nAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsnValProValGlyPr 425
Db 1128 CAACCCAACTGATGGAACATGGACATTAAGTTGTAAAGCTACAGC-----GGAAG 1178
Qy 425 cGlnThrPheSerLeuAlaIleValAla 434
Db 1179 TGCAACTATCAAGTAGATGCTGAAGT 1206

RESULT 9
US-08-894-818B-34
Sequence 34, Application US/08994818B
Patent No. 6261822
GENERAL INFORMATION:
APPLICANT: TAKAKURA, Hikaru
APPLICANT: MORISHITA, Mio
APPLICANT: YAMAMOTO, Katsuhiko
APPLICANT: MITTA, Masanori
APPLICANT: ASADA, Kiyozo
APPLICANT: TSUNASAWA, Susumu
APPLICANT: KATO, Ikunoshin
TITLE OF INVENTION: HYPERTHERMOSTABLE PROTEASE GENES
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
ADDRESSEE: Browdy and Neimark
STREET: 419 Seventh Street N.W., Ste. 300
CITY: Washington
STATE: D.C.
COUNTRY: United States of America
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/894,818B
FILING DATE: 20-MAY-1998
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP96/03253
FILING DATE: 07-NOV-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 323285/1995
FILING DATE: 12-DEC-1995
ATTORNEY/AGENT INFORMATION:
NAME: Browdy, Roger L.
REGISTRATION NUMBER: 25,618
REFERENCE/DOCKET NUMBER: TAKAKURA=1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 628-5197
TELEFAX: (202) 737-3528
INFORMATION FOR SEQ ID NO: 34:
SEQUENCE CHARACTERISTICS:
LENGTH: 1962 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: genomic DNA
US-08-894-818B-34
Alignment Scores:
Pred. No.: 4,69e-33 Length: 1962
Score: 411.50 Matches: 140
Percent Similarity: 42.55% Conservative: 60
Best Local Similarity: 29.79% Mismatches: 148
Query Match: 18.31% Indels: 122
DB: 3 Gaps: 20
US-09-985-689A-1-COPY (1-434) x US-08-894-818B-34 (1-1962)
Qy 12 ValAlaGlnSerSerTyrGly-LeuTyr-----GlyGlnGlyG1 24
Db 420 GTCTGCAGCTCAAGTTATGGCAACTTACGTTTGGAACTGGGATATGATGTTCTGAAT 479
Qy 24 nIleValAlaAlaAspThrGlyLeuAspThrGlyArgAsnAspSerMetHisG1 44
Db 480 CACAATAGGAATAATTGACACTGGCAATTGAC-----GTTCTCATCC 521
Qy 44 uAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsnAsnAlaAen-- 63
Db 522 AGATCTCCAGGAAGAAGTA-----ATTGGGTGGGTAGATTTTGTCAATGG 566
Qy 64 -----AspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuG1 78
Db 567 TAGGAGTTATCCATACGATGACCATGGACATGGAACTCATGTAGCTTCAATAGCAGCTGG 626
Qy 78 yAsnGlySerThrAsn-----LysGlyMetAlaProGlnAlaAsnLeuValPh 94
Db 627 TACTGGAGCAGCAAGTAATGCAAGTACAGGGAATGCTCCAGAGCTAAGCTGGGGGG 686
Qy 94 eGlnSerIleMet-----AspSerGlyGlyGlyLeuGlyGlyLeuProSerAsnLeuG1 112
Db 687 AATTAAGGTTCTAGTGCGCGATGGTCTTGGAGCATATCTACTATATTAAGGGAGTTGA 746
Qy 112 nThrLeuPheSerGlnAlaTyrSerAlaGlyAlaArgIleHisThrAsnSerTrpGlyAl 132
Db 747 GTGGCCCGTTGATAACAAAGATAGTACGGAATTAAGTCAATTAATCTTCTCTGTGTC 806
Qy 132 a-----AlaValAsnGlyAlaTyrThrTh 140
Db 807 AAGCCAGAGCTCAGATGATGCTAGCTGCTTAACTAGTCAGCTGTTATTCAGCAGCTGGGATGC 866
Qy 140 rAspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAl 160
Db 867 T-----GGATTAGTTGTGTGTGTCGCGC 890
Qy 160 aGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaI1 180

891 TGAACACAGTGGACCTAACAGATACAAATCGTTCTCCAGCAGCTGCAAGCAAGTTAT 950
180 eThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerThrAlaAspAsnIleAs 200
951 TACAGTTGGAGCC-----GTTGCAAGTATGA 977
200 nHisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVa 220
978 TGTATTAAACAGCTTCTCAAGCAGAGGGCCAACTGACAGCGCAGCTTAAGCCTGAGGT 1037
220 lMetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPh 240
1038 TGTCTCTCAGGAATATGATTAATGCTGCCAGCAAGT-----GGAAGTATGAT 1088
240 eTrpAlaAsnHisAspSerLysThrAlaThrMetGlyGlyThrSerMetAlaThrProIl 260
1089 GGGTCAACCAATTAATGACTATTACAGCAGCAGCTCTGGCAGCATCAATGGCACTCTCTCA 1148
260 eValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPr 280
1149 CGTAGCTGTATTGCGAGCCCTCTGCTCCAA-----GCACACCC 1187
280 oLys-----ProSerLeuLeuLysAlaAlaLeuIleAlaGlyAla----- 293
1188 GAGCTGAGCTCCAGCAAAAGTAAACAGCCCTCATAGAACTGCTGATATCGTAAAGCC 1247
294 -----AlaAspIleGlyLeuGlyTyrProAsnGlyAsnGlnGlyTyrGlyArgVa 310
1248 AGATGAATAGCGATATAGCTTACGCTGCA-----GCTAGGGT 1286
310 lThrLeuAspLysSerLeuAsnValAlaThrValAsnGluSerSerSerLeuSerThrSe 330
1287 TAATGATATACAGCTATTAAC-----TACGATTAATATGCAAGCTAGTCTCACTGG 1340
330 rGlnLysAla-----ThrTyrSerPheThrAlaThrAlaGlyLysProLe 345
1341 ATATGTTGCCAACAAAGGCGAGCAAACTCACCAGTTCTGTTATTAGCGGAGCTTCGTCGT 1400
345 uLysIleSerLeuValTrpSerAspAlaProLaserThrThrAlaSerValThrLeuVa 365
1401 AACTGCCATATATCTGGCAATGCCAAT----- 1431
365 lAsnAspLeuAsnLeuValIleThrAlaProAsnGlyThrGlnTyrValGlyAsnAspPh 385
1432 -ACGACCTTGATCTTTACTCTACGATCCCAATGGAAACAG--GTTGACTACTCTTA 1487
385 eThrSerProTyrAsnAspAsnTrpAspGlyArgAsnAsnValGluAsnValPheIleAs 405
1488 CACCGCTACTAT-----GGATTGCAAAAGGTTGGTTATTA 1523
405 nAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaThrAsnValProValGlyPr 425
1524 CAACCCAACTGATGGAACATGGCAATTAAGGTTGTAAGCTACAGC-----GGAAG 1574
425 oGlnThrPheSerLeuAlaIleValAsn 434
1575 TGCAAACTATCAAGTAGATGTGTAAAGT 1602

RESULT 10

US-09-445-472-15
; Sequence 15, Application US/09445472
; Patent No. 6358726
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: SHIMOJO, Tomoko
; APPLICANT: ASADA, Kiyozo
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
; FILE REFERENCE: TAKAKURA=6
; CURRENT APPLICATION NUMBER: US/09/445,472
; CURRENT FILING DATE: 1999-12-06

; PRIOR APPLICATION NUMBER: 151969/1997
; PRIOR FILING DATE: 1997-06-10
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 15
; LENGTH: 1962
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
US-09-445-472-15

Alignment Scores:
Pred. No.: 4 696-33 Length: 1962
Score: 411.50 Matches: 140
Percent Similarity: 42.55% Conservative: 60
Best Local Similarity: 29.79% Mismatches: 148
Query Match: 18.31% Indels: 122
DB: 4 Gaps: 20

US-09-985-689A-1-COPY (1-434) x US-09-445-472-15 (1-1962)

Qy 12 ValAlaGlnSerSerTyrGly-LeuTyr-----GlyGlnGlyGl 24
Db 420 GTCTGAGCTCAAGTTATGCGCAACTTACGTTGGAACTGGGATATGATGGTTCTCGAAT 479
Qy 24 nIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisGl 44
Db 480 CACATAGGATTAATGACACTGGATTCAC-----GCTTCTCATCC 521
Qy 44 uAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsnAsnAlaAsn-- 63
Db 522 AGATCTCCAGGAAAGTA-----ATTGGGTGGGTAGATTTGTCAATGG 566
Qy 64 -----AspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGl 78
Db 567 TAGAGTTATCCATACGATGACCATGGACATGGAACTCATGTAGCTTCAATAGCAGCTGG 626
Qy 78 YAsnGlySerThrAsn-----LysGlyMetAlaProGlnAlaAsnLeuValPh 94
Db 627 TACTGGAGACCAAGTAATGGCAAGTACAGGGAATGGCTCCAGGAGCTTAAGCTGCGGG 686
Qy 94 eGlnSerIleMet-----AspSerGlyGlyGlyLeuGlyGlyLeuProSerAsnLeuGl 112
Db 687 AATTAAGTTCTAGTCCGATGGTTCTGGAGCATATCTACTATTAATTAAGGAGTTGA 746
Qy 112 nThrLeuPheSerGlnAlaThrSerAlaGlyAlaArgIleHisThrAsnSerTrpGlyAl 132
Db 747 GTGGCCGCTTTCATAACAAAGATAAGTACGGAATTAAGTTCATTAATCTTTCTCTGTTTC 806
Qy 132 a-----AlaValAsnGlyAlaTyrThrTh 140
Db 807 AAGCCAGAGCTCAGATGGTACTGACGCTCTAAGTCAGGCTGTTAATGACAGCTGGGATGC 866
Qy 140 rAspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAl 160
Db 867 T-----GGATTAGTTCTGTTGGTGGCGC 890
Qy 160 aGlyAsnGluGlyProAsnGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIl 180
Db 891 TGGAAACAGTGGACCTAACCAAGTATACATCGGTTCTCCAGCAGCTCACAAGCAAGTTAT 950
Qy 180 eThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerThrAlaAspAsnIleAs 200
Db 951 TACAGTTGGAGCC-----GTTGCAAGTATGA 977
Qy 200 nHisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVa 220
Db 978 TGTATTAAACAGCTTCTCAAGCAGAGGGCCAACTGACAGCGCAGCTTAAGCCTGAGGT 1037
Qy 220 lMetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPh 240
Db 1038 TGTGCTCCAGAACTGGATTAATGCTGCCAGAGCAGT-----GGAAGTATGAT 1088

```
QY 240 eTpaAAsnHiaSerpSerlyTyraTyMetGlyGlyThrSerMetAlaThrPro1 260
Db 1089 GGGTCAACCAATTAATGACTATTACACAGCAGCTCTCGACATCAATGCCAACTCTCA 1148
QY 260 eValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyLeuThrPr 280
Db 1149 CGTAGCTGGTATTGCGACCCCTCTTGCTCAA-----GCACACCC 1187
QY 280 Olys-----ProSerLeuLeuLysAlaLeuLeuAlaGlyAla----- 293
Db 1188 GAGCTGGACTCCAGACAAAGTAAACAGCCCTCATAGAACTGCTGATATCGTAAAGCC 1247
QY 294 -----AlaAspLeuGlyLeuGlyTyProAsnGlyAsnGlnGlyTyProGlyArgVa 310
Db 1248 AGATGAATAGCCGATATAGCCCTACCGTGCA-----GGTAGGCT 1286
QY 310 lThrLeuAspLysSerLeuAsnValAlaTyValAsnGluSerSerSerLeuSerThrSe 330
Db 1287 TAATGCATACAGGCTATAAC-----TACGATACTATGCAAGCTAGTGTCACTGG 1340
QY 330 rGlnLysAla-----ThrTySerPheThrAlaThrAlaGlyLysProLe 345
Db 1341 ATATGTTGCCAACAAGGCGCAAACTCACCAGTTGCTTATTAGCGGAGCTTCGTTCGT 1400
QY 345 uLysLeuSerLeuValTrpSerAspAlaProAlaSerThrThrAlaSerValThrLeuVa 365
Db 1401 AACTGCCACATTAATGCGACATGCAAT----- 1431
QY 365 lAsnAspLeuAsnLeuValLeuThrAlaProAsnGlyThrGlnTyValGlyAsnAspPh 385
Db 1432 -AGCGACCTTGATCTTTACCTCTACGATCCCAATGGAACACAG---GTTGACTACTCTTA 1487
QY 385 eThrSerProTyAsnAspAsnTrpAspGlyArgAsnAsnValGluAsnValPheLeAs 405
Db 1488 CACGCGCTACTAT-----GGATTCGAAAGGTTGTTATTA 1523
QY 405 nAlaProGlnSerGlyThrTyThrileGluValGlnAlaTyAsnValProValGlyPr 425
Db 1524 CAACCCAACTGATGGAACATGACAAATAAGTTGTAAGCTACAGC-----GGAAG 1574
QY 425 oGlnThrPheSerLeuAlaLeuValAsn 434
Db 1575 TGCAAACTATCAAGTAGATGGTAAGT 1602
```

RESULT 11

```
US-08-894-818B-6
; Sequence 6, Application US/08894818B
; Patent No. 6261822
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: YAMAMOTO, Katsuhiko
; APPLICANT: MITTA, Masanori
; APPLICANT: ASADA, Kiyozo
; APPLICANT: TSUNASAWA, Susumu
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: HYPERTERMOSTABLE PROTEASE GENES
; NUMBER OF SEQUENCES: 42
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Browdy and Neimark
; STREET: 419 Seventh Street N.W., Ste. 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: United States of America
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/894,818B
```

```
; FILING DATE: 20-MAY-1998
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/JP96/03253
; FILING DATE: 07-NOV-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 323285/1995
; FILING DATE: 12-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Browdy, Roger L.
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: TAKAKURA=1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 628-5197
; TELEFAX: (202) 737-3528
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1977 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
; US-08-894-818B-6
```

Alignment Scores:

```
Pred. No.: 1,08e-31 Length: 1977
Score: 398.50 Matches: 133
Percent Similarity: 43.07% Conservative: 69
Best Local Similarity: 28.36% Mismatches: 147
Query Match: 17.73% Indels: 120
DB: 3 Gaps: 20
```

US-09-985-689A-1-COPY (1-434) x US-08-894-818B-6 (1-1977)

```
QY 12 ValAlaGlnSerSerTyGly-LeuTy-----GlyGlnGly 24
Db 420 GTCCTCAGCTCAAGTATATGGCAACTTACGTTGGACCTGGATATGATGTTCTGGAAT 479
QY 24 nLeValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisG1 44
Db 480 CACAATAGGAATAATTGACACTGGAAATTGAC-----GCTTCTCATCC 521
QY 44 uAlaPheArgGlyLysIleThrAlaLeuTyAlaLeuGlyArgThrAsnAsnAlaAsn-- 63
Db 522 AGATCTCCAAGAAAAGTA-----ATTGGTGGGTAGATTTTGTCAATGG 566
QY 64 -----AspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuG1 78
Db 567 TAGGAGTTATCCATACGATGACCATGGACATGGAACATCATGTAGCTTCAATAGCAGCTGG 626
QY 78 yAsnGlySerThrAsn-----LysGlyMetAlaProGlnAlaAsnLeuValPh 94
Db 627 TACTGGACGACGACGATATGGCAAGTACAGGGAATGCTCCAGAGCTAAGCTGCGCGG 686
QY 94 eGlnSerIleMet-----AspSerGlyGlyLeuGlyLeuProSerAsnLeuG1 112
Db 687 AATTAAGGTTCTAGGTGCGGATGTTCTGGAAGCATCTACTATAATAATTAAGGGAGTTGA 746
QY 112 nThrLeuPheSerGlnAlaTySerAlaGlyAlaArgile----- 125
Db 747 GTGGGCGGTGATAACAAAGATAAGTACGGAATTAAGGTCAATTAATCTTCTCTTGGTTC 806
QY 126 -----HisThrAsnSerTrpGlyAlaAlaValAsnGlyAlaTyThrTh 140
Db 807 AAGCCAGAGCTCCGACGGAGCCGACTCCCTCAGTCAGCGCGTCAACACGCGCTGGACGC 866
QY 140 rAspSerArgAsnValAspAspTyValArgLysAsnAspMetThrIleLeuPheAlaAl 160
Db 867 C-----GGTATAGTAGTCTGCGTCGCGCGC 890
QY 160 aGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAla1 180
Db 891 CGGCAACACGCGGCGCGAACACCTACACCGTCGCTCGCTCCCGCGCGGAGCAAGTCTAT 950
```


180 eThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerThrAlaAspAsnIleAs 200
181 |||||
182 |||||
183 |||||
184 |||||
185 |||||
186 |||||
187 |||||
188 |||||
189 |||||
190 |||||
191 |||||
192 |||||
193 |||||
194 |||||
195 |||||
196 |||||
197 |||||
198 |||||
199 |||||
200 nHisValAlaGlnPheSerArgGlyProThrLysAspGlyArgIleLysProAspVa 220
201 |||||
202 |||||
203 |||||
204 |||||
205 |||||
206 |||||
207 |||||
208 |||||
209 |||||
210 |||||
211 |||||
212 |||||
213 |||||
214 |||||
215 |||||
216 |||||
217 |||||
218 |||||
219 |||||
220 lMetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerPh 240
221 |||||
222 |||||
223 |||||
224 |||||
225 |||||
226 |||||
227 |||||
228 |||||
229 |||||
230 |||||
231 |||||
232 |||||
233 |||||
234 |||||
235 |||||
236 |||||
237 |||||
238 |||||
239 |||||
240 eTrpAlaAsnHisAspSerLysAlaThrMetGlyThrSerMetAlaThrProil 260
241 |||||
242 |||||
243 |||||
244 |||||
245 |||||
246 |||||
247 |||||
248 |||||
249 |||||
250 |||||
251 |||||
252 |||||
253 |||||
254 |||||
255 |||||
256 |||||
257 |||||
258 |||||
259 |||||
260 |||||
261 |||||
262 |||||
263 |||||
264 |||||
265 |||||
266 eValAlaGly---AsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleTh 279
267 |||||
268 |||||
269 |||||
270 |||||
271 |||||
272 |||||
273 |||||
274 |||||
275 |||||
276 |||||
277 |||||
278 |||||
279 |||||
280 |||||
281 |||||
282 |||||
283 |||||
284 |||||
285 |||||
286 |||||
287 |||||
288 |||||
289 |||||
290 |||||
291 |||||
292 |||||
293 rProLys-----ProSerLeuLeuLysAlaAlaLeuIleAlaGlyAla----- 293
294 |||||
295 |||||
296 |||||
297 |||||
298 |||||
299 |||||
300 |||||
301 |||||
302 |||||
303 |||||
304 |||||
305 |||||
306 |||||
307 |||||
308 |||||
309 |||||
310 |||||
311 |||||
312 |||||
313 |||||
314 |||||
315 |||||
316 |||||
317 |||||
318 |||||
319 |||||
320 |||||
321 |||||
322 |||||
323 |||||
324 |||||
325 |||||
326 rLeuSerThrSerGlnLysAlaThrThrSerPheThrAlaThrAlaGlyLysProLeuLys 346
327 |||||
328 |||||
329 |||||
330 |||||
331 |||||
332 |||||
333 |||||
334 |||||
335 |||||
336 |||||
337 |||||
338 |||||
339 |||||
340 |||||
341 |||||
342 |||||
343 |||||
344 |||||
345 |||||
346 |||||
347 |||||
348 |||||
349 |||||
350 |||||
351 |||||
352 |||||
353 |||||
354 |||||
355 |||||
356 |||||
357 |||||
358 |||||
359 |||||
360 |||||
361 |||||
362 |||||
363 |||||
364 |||||
365 |||||
366 nAspLeuAsnLeuValIleThrAlaProAsnGlyThrGlnThrValGlyAsnAspPheTh 386
367 |||||
368 |||||
369 |||||
370 |||||
371 |||||
372 |||||
373 |||||
374 |||||
375 |||||
376 |||||
377 |||||
378 |||||
379 |||||
380 |||||
381 |||||
382 |||||
383 |||||
384 |||||
385 |||||
386 |||||
387 |||||
388 |||||
389 |||||
390 |||||
391 |||||
392 |||||
393 |||||
394 |||||
395 |||||
396 |||||
397 |||||
398 |||||
399 |||||
400 |||||
401 |||||
402 |||||
403 |||||
404 |||||
405 |||||
406 aProGlnSerGlyThrThrThrIleGluValGlnAlaThrAsnValProValGlyProGl 426
407 |||||
408 |||||
409 |||||
410 |||||
411 |||||
412 |||||
413 |||||
414 |||||
415 |||||
416 |||||
417 |||||
418 |||||
419 |||||
420 |||||
421 |||||
422 |||||
423 |||||
424 |||||
425 |||||
426 nThrPheSerLeuAlaIleValAsn 434
427 |||||
428 |||||
429 |||||
430 |||||
431 |||||
432 |||||
433 |||||
434 |||||
435 |||||
436 |||||
437 |||||
438 |||||
439 |||||
440 |||||
441 |||||
442 |||||
443 |||||
444 |||||
445 |||||
446 |||||
447 |||||
448 |||||
449 |||||
450 |||||
451 |||||
452 |||||
453 |||||
454 |||||
455 |||||
456 |||||
457 |||||
458 |||||
459 |||||
460 |||||
461 |||||
462 |||||
463 |||||
464 |||||
465 |||||
466 |||||
467 |||||
468 |||||
469 |||||
470 |||||
471 |||||
472 |||||
473 |||||
474 |||||
475 |||||
476 |||||
477 |||||
478 |||||
479 |||||
480 |||||
481 |||||
482 |||||
483 |||||
484 |||||
485 |||||
486 |||||
487 |||||
488 |||||
489 |||||
490 |||||
491 |||||
492 |||||
493 |||||
494 |||||
495 |||||
496 |||||
497 |||||
498 |||||
499 |||||
500 |||||
501 |||||
502 |||||
503 |||||
504 |||||
505 |||||
506 |||||
507 |||||
508 |||||
509 |||||
510 |||||
511 |||||
512 |||||
513 |||||
514 |||||
515 |||||
516 |||||
517 |||||
518 |||||
519 |||||
520 |||||
521 |||||
522 |||||
523 |||||
524 |||||
525 |||||
526 |||||
527 |||||
528 |||||
529 |||||
530 |||||
531 |||||
532 |||||
533 |||||
534 |||||
535 |||||
536 |||||
537 |||||
538 |||||
539 |||||
540 |||||
541 |||||
542 |||||
543 |||||
544 |||||
545 |||||
546 |||||
547 |||||
548 |||||
549 |||||
550 |||||
551 |||||
552 |||||
553 |||||
554 |||||
555 |||||
556 |||||
557 |||||
558 |||||
559 |||||
560 |||||
561 |||||
562 |||||
563 |||||
564 |||||
565 |||||
566 |||||
567 |||||
568 |||||
569 |||||
570 |||||
571 |||||
572 |||||
573 |||||
574 |||||
575 |||||
576 |||||
577 |||||
578 |||||
579 |||||
580 |||||
581 |||||
582 |||||
583 |||||
584 |||||
585 |||||
586 |||||
587 |||||
588 |||||
589 |||||
590 |||||
591 |||||
592 |||||
593 |||||
594 |||||
595 |||||
596 |||||
597 |||||
598 |||||
599 |||||
600 |||||
601 |||||
602 |||||
603 |||||
604 |||||
605 |||||
606 |||||
607 |||||
608 |||||
609 |||||
610 |||||
611 |||||
612 |||||
613 |||||
614 |||||
615 |||||
616 |||||
617 |||||
618 |||||
619 |||||
620 |||||
621 |||||
622 |||||
623 |||||
624 |||||
625 |||||
626 |||||
627 |||||
628 |||||
629 |||||
630 |||||
631 |||||
632 |||||
633 |||||
634 |||||
635 |||||
636 |||||
637 |||||
638 |||||
639 |||||
640 |||||
641 |||||
642 |||||
643 |||||
644 |||||
645 |||||
646 |||||
647 |||||
648 |||||
649 |||||
650 |||||
651 |||||
652 |||||
653 |||||
654 |||||
655 |||||
656 |||||
657 |||||
658 |||||
659 |||||
660 |||||
661 |||||
662 |||||
663 |||||
664 |||||
665 |||||
666 |||||
667 |||||
668 |||||
669 |||||
670 |||||
671 |||||
672 |||||
673 |||||
674 |||||
675 |||||
676 |||||
677 |||||
678 |||||
679 |||||
680 |||||
681 |||||
682 |||||
683 |||||
684 |||||
685 |||||
686 |||||
687 |||||
688 |||||
689 |||||
690 |||||
691 |||||
692 |||||
693 |||||
694 |||||
695 |||||
696 |||||
697 |||||
698 |||||
699 |||||
700 |||||
701 |||||
702 |||||
703 |||||
704 |||||
705 |||||
706 |||||
707 |||||
708 |||||
709 |||||
710 |||||
711 |||||
712 |||||
713 |||||
714 |||||
715 |||||
716 |||||
717 |||||
718 |||||
719 |||||
720 |||||
721 |||||
722 |||||
723 |||||
724 |||||
725 |||||
726 |||||
727 |||||
728 |||||
729 |||||
730 |||||
731 |||||
732 |||||
733 |||||
734 |||||
735 |||||
736 |||||
737 |||||
738 |||||
739 |||||
740 |||||
741 |||||
742 |||||
743 |||||
744 |||||
745 |||||
746 |||||
747 |||||
748 |||||
749 |||||
750 |||||
751 |||||
752 |||||
753 |||||
754 |||||
755 |||||
756 |||||
757 |||||
758 |||||
759 |||||
760 |||||
761 |||||
762 |||||
763 |||||
764 |||||
765 |||||
766 |||||
767 |||||
768 |||||
769 |||||
770 |||||
771 |||||
772 |||||
773 |||||
774 |||||
775 |||||
776 |||||
777 |||||
778 |||||
779 |||||
780 |||||
781 |||||
782 |||||
783 |||||
784 |||||
785 |||||
786 |||||
787 |||||
788 |||||
789 |||||
790 |||||
791 |||||
792 |||||
793 |||||
794 |||||
795 |||||
796 |||||
797 |||||
798 |||||
799 |||||
800 |||||
801 |||||
802 |||||
803 |||||
804 |||||
805 |||||
806 |||||
807 |||||
808 |||||
809 |||||
810 |||||
811 |||||
812 |||||
813 |||||
814 |||||
815 |||||
816 |||||
817 |||||
818 |||||
819 |||||
820 |||||
821 |||||
822 |||||
823 |||||
824 |||||
825 |||||
826 |||||
827 |||||
828 |||||
829 |||||
830 |||||
831 |||||
832 |||||
833 |||||
834 |||||
835 |||||
836 |||||
837 |||||
838 |||||
839 |||||
840 |||||
841 |||||
842 |||||
843 |||||
844 |||||
845 |||||
846 |||||
847 |||||
848 |||||
849 |||||
850 |||||
851 |||||
852 |||||
853 |||||
854 |||||
855 |||||
856 |||||
857 |||||
858 |||||
859 |||||
860 |||||
861 |||||
862 |||||
863 |||||
864 |||||
865 |||||
866 |||||
867 |||||
868 |||||
869 |||||
870 |||||
871 |||||
872 |||||
873 |||||
874 |||||
875 |||||
876 |||||
877 |||||
878 |||||
879 |||||
880 |||||
881 |||||
882 |||||
883 |||||
884 |||||
885 |||||
886 |||||
887 |||||
888 |||||
889 |||||
890 |||||
891 |||||
892 |||||
893 |||||
894 |||||
895 |||||
896 |||||
897 |||||
898 |||||
899 |||||
900 |||||
901 |||||
902 |||||
903 |||||
904 |||||
905 |||||
906 |||||
907 |||||
908 |||||
909 |||||
910 |||||
911 |||||
912 |||||
913 |||||
914 |||||
915 |||||
916 |||||
917 |||||
918 |||||
919 |||||
920 |||||
921 |||||
922 |||||
923 |||||
924 |||||
925 |||||
926 |||||
927 |||||
928 |||||
929 |||||
930 |||||
931 |||||
932 |||||
933 |||||
934 |||||
935 |||||
936 |||||
937 |||||
938 |||||
939 |||||
940 |||||
941 |||||
942 |||||
943 |||||
944 |||||
945 |||||
946 |||||
947 |||||
948 |||||
949 |||||
950 |||||
951 |||||
952 |||||
953 |||||
954 |||||
955 |||||
956 |||||
957 |||||
958 |||||
959 |||||
960 |||||
961 |||||
962 |||||
963 |||||
964 |||||
965 |||||
966 |||||
967 |||||
968 |||||
969 |||||
970 |||||
971 |||||
972 |||||
973 |||||
974 |||||
975 |||||
976 |||||
977 |||||
978 |||||
979 |||||
980 |||||
981 |||||
982 |||||
983 |||||
984 |||||
985 |||||
986 |||||
987 |||||
988 |||||
989 |||||
990 |||||
991 |||||
992 |||||
993 |||||
994 |||||
995 |||||
996 |||||
997 |||||
998 |||||
999 |||||
1000 |||||

ZIP: 20006
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
COMPUTER: IBM Compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/000,016
FILING DATE: January 30, 1998
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Warren M. Cheek, Jr.
REGISTRATION NUMBER: 33,367
REFERENCE/DOCKET NUMBER:
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-721-8200
TELEFAX: 202-721-8250
TELEX:
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 2539 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: Genomic DNA
ORIGINAL SOURCE:
ORGANISM: Streptomyces viridosporus
STRAIN: A-914
NAME/KEY: CDS
LOCATION: 338...2539
IDENTIFICATION METHOD: E
US-09-000-016-3
Alignment Scores:
Pred. No.: 9,1e-27 Length: 2539
Score: 353.00 Matches: 131
Percent Similarity: 42.63% Conservative: 54
Best Local Similarity: 30.18% Mismatches: 159
Query Match: 15.71% Indels: 90
DB: 3 Gaps: 16
US-09-985-689A-1-COPY (1-434) x US-09-000-016-3 (1-2539)
QY 2 AspValAlaArgGlyIleValLysAlaAspValAlaGlnSerSerThrGlyLeuThrGly 21
Db 953 GACAGTCCTCGCGGAGATCGCGCCCAAGCGGTGTCGCGC---GGCTACGACGGC 1009
QY 22 GlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 41
Db 1010 AAGGCGGTGAAGATCGCGCTCTGACACCGGTGTCGACACGAGC----- 1054
QY 42 MethHisGluAlaPheArgGlyLysIleThrAlaLeuThrAlaLeuGlyArgThrAsn 61
Db 1055 ---CATCCGACCTGAGGCGCGGTGACCGCTCCAGAGACTTCACCGCGCGCGCGC 1111
QY 62 AlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 81
Db 1112 GCCGCGCAAGGTGGCGCACCGCACCGCTCGATCGCGCGCGCGCGCGCGC 1171
QY 82 ThrAsn-----LysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIle 97
Db 1172 CAGTCCAAAGGGCAAGTACAGGGGTGCGCACCCCGCGCGCGCGCGCGCGCGC 1231
QY 98 MetAspSerGlyGlyLeuGlyLeuProSerAsnLeuGlnThrLeuPheSerGln 117
Db 1232 CTCGACGACTCCGT-----TTGCGGACGACTCCGCGCATCTCCGCGCATGGAGTGG 1285
QY 118 AlaTyrSerAlaGlyAlaArgIleHisThrAsnSerThrGlyAlaAlaValAsnGly 137
Db 118 |||||
119 |||||
120 |||||
121 |||||
122 |||||
123 |||||
124 |||||
125 |||||
126 |||||
127 |||||
128 |||||
129 |||||
130 |||||
131 |||||
132 |||||
133 |||||
134 |||||
135 |||||
136 |||||
137 |||||
138 |||||
139 |||||
140 |||||
141 |||||
142 |||||
143 |||||
144 |||||
145 |||||
146 |||||
147 |||||
148 |||||
149 |||||
150 |||||
151 |||||
152 |||||
153 |||||
154 |||||
155 |||||
156 |||||
157 |||||
158 |||||
159 |||||
160 |||||
161 |||||
162 |||||
163 |||||
164 |||||
165 |||||
166 |||||
167 |||||
168 |||||
169 |||||
170 |||||
171 |||||
172 |||||
173 |||||
174 |||||
175 |||||
176 |||||
177 |||||
178 |||||
179 |||||
180 |||||
181 |||||
182 |||||
183 |||||
184 |||||
185 |||||
186 |||||
187 |||||
188 |||||
189 |||||
190 |||||
191 |||||
192 |||||
193 |||||
194 |||||
195 |||||
196 |||||
197 |||||
198 |||||
199 |||||
200 |||||
201 |||||
202 |||||
203 |||||
204 |||||
205 |||||
206 |||||
207 |||||
208 |||||
209 |||||
210 |||||
211 |||||
212 |||||
213 |||||
214 |||||
215 |||||
216 |||||
217 |||||
218 |||||
219 |||||
220 |||||
221 |||||
222 |||||
223 |||||
224 |||||
225 |||||
226 |||||
227 |||||
228 |||||
229 |||||
230 |||||
231 |||||
232 |||||
233 |||||
234 |||||
235 |||||
236 |||||
237 |||||
238 |||||
239 |||||
240 |||||
241 |||||
242 |||||
243 |||||
244 |||||
245 |||||
246 |||||
247 |||||
248 |||||
249 |||||
250 |||||
251 |||||
252 |||||
253 |||||
254 |||||
255 |||||
256 |||||
257 |||||
258 |||||
259 |||||
260 |||||
261 |||||
262 |||||
263 |||||
264 |||||
265 |||||
266 |||||
267 |||||
268 |||||
269 |||||
270 |||||
271 |||||
272 |||||
273 |||||
274 |||||
275 |||||
276 |||||
277 |||||
278 |||||
279 |||||
280 |||||
281 |||||
282 |||||
283 |||||
284 |||||
285 |||||
286 |||||
287 |||||
288 |||||
289 |||||
290 |||||
291 |||||
292 |||||
293 |||||
294 |||||
295 |||||
296 |||||
297 |||||
298 |||||
299 |||||
300 |||||
301 |||||
302 |||||
303 |||||
304 |||||
305 |||||
306 |||||
307 |||||
308 |||||
309 |||||
310 |||||
311 |||||
312 |||||
313 |||||
314 |||||
315 |||||
316 |||||
317 |||||
318 |||||
319 |||||
320 |||||
321 |||||
322 |||||
323 |||||
324 |||||
325 |||||
326 |||||
327 |||||
328 |||||
329 |||||
330 |||||
331 |||||
332 |||||
333 |||||
334 |||||
335 |||||
336 |||||
337 |||||
338 |||||
339 |||||
340 |||||
341 |||||
342 |||||
343 |||||
344 |||||
345 |||||
346 |||||
347 |||||
348 |||||
349 |||||
350 |||||
351 |||||
352 |||||
353 |||||
354 |||||
355 |||||
356 |||||
357 |||||
358 |||||
359 |||||
360 |||||
361 |||||
362 |||||
363 |||||
364 |||||
365 |||||
366 |||||
367 |||||
368 |||||
369 |||||
370 |||||
371 |||||
372 |||||
373 |||||
374 |||||
375 |||||
376 |||||
377 |||||
378 |||||
379 |||||
380 |||||
381 |||||
382 |||||
383 |||||
384 |||||
385 |||||
386 |||||
387 |||||
388 |||||
389 |||||
390 |||||
391 |||||
392 |||||
393 |||||
394 |||||
395 |||||
396 |||||
397 |||||
398 |||||
399 |||||
400 |||||
401 |||||
402 |||||
403 |||||
404 |||||
405 |||||
406 |||||
407 |||||
408 |||||
409 |||||
410 |||||
411 |||||
412 |||||
413 |||||
414 |||||
415 |||||
416 |||||
417 |||||
418 |||||
419 |||||
420 |||||
421 |||||
422 |||||
423 |||||
424 |||||
425 |||||
426 |||||
427 |||||
428 |||||
429 |||||
430 |||||
431 |||||
432 |||||
433 |||||
434 |||||
435 |||||
436 |||||
437 |||||
438 |||||
439 |||||
440 |||||
441 |||||
442 |||||
443 |||||
444 |||||
445 |||||
446 |||||
447 |||||
448 |||||
449 |||||
450 |||||
451 |||||
452 |||||
453 |||||
454 |||||
455 |||||
456 |||||
457 |||||
458 |||||
459 |||||
460 |||||
461 |||||
462 |||||
463 |||||
464 |||||
465 |||||
466 |||||
467 |||||
468 |||||
469 |||||
470 |||||
471 |||||
472 |||||
473 |||||
474 |||||
475 |||||
476 |||||
477 |||||
478 |||||
479 |||||
480 |||||
481 |||||
482 |||||
483 |||||
484 |||||
485 |||||
486 |||||
487 |||||
488 |||||
489 |||||
490 |||||
491 |||||
492 |||||
493 |||||
494 |||||
495 |||||
496 |||||
497 |||||
498 |||||
499 |||||
500 |||||
501 |||||
502 ||


```

QY 98 MetAspSerGlyGlyLeuGlyLeuProSerAsnLeuGlnThrLeuPheSerGln 117
Db 1232 CTCGACGACTCGGT-----TTGGCGACACTCCGGCATCTCCGGCGCATGGATGG 1285
QY 118 AlaTyrSerAlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAsnGlyAla 137
Db 1286 GCGCGCGCGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1333
QY 138 TyrThrThrAspSerArgAsnValAspAspTyrValArgLys-----AsnAspMetThr 155
Db 1334 GACACACCGGAGACGACCGCTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1393
QY 156 IleLeuPhe-----AlaAlaGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaPro 173
Db 1394 GTCTGTTCGCCATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1447
QY 174 GlyThrAlaLysAsnAlaIleThrValGlyAlaThrGluAsnLeuArgProSerPheGly 193
Db 1448 GGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1480
QY 194 SerTyrAlaAspAsnIleAsnHisValAlaGlnPheSerSerArgGlyPro---ThrIys 212
Db 1481 -----GTCAGCAGACAGACAGCTCCGCGACTTCTCTCCACGCGCGCGCGCGCTCGCG 1534
QY 213 AspGlyArgIleLysProAspValMetAlaProGlyThrPheIleLeuSerAlaArgSer 232
Db 1535 GACGCGCGCATCAAGCGCGCATCACCCTCCGCGCGTGGACATCAGCGCGCGCGCTCGCGG 1594
QY 233 SerLeuAlaProAspSerSerPheTrpAlaAsnHisAspSerLysTyrAlaTyrMetGly 252
Db 1595 GAGGCAACGACATCGCGCGAGAGTGTGTGAGGAGCGCGCGCGCGCTACATGACCATCTCC 1654
QY 253 GlyThrSerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPhe 272
Db 1655 GGCAGCTGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1711
QY 273 ValLysAsnArgGlyIleThrProLysProSerLeuLeuLysAlaAlaLeu----- 289
Db 1712 -----CACCCGACTCGACTCCGCGCGACTCGAGGCGCG 1747
QY 290 IleAlaGlyAlaAlaAspIleGlyLeuGlyTyrProAsnGlyAsnGlnGlyTyrGlyArg 309
Db 1748 CTCACGCGCTCCACCAAG---GGCGGCAAGTACACCCCGTTTCGAGCAGGGTTCGGCGCGG 1804
QY 310 ValThrLeuAspLysSerLeuAsnValAlaTyrValAsnGluSerSerSerLeuSer--- 328
Db 1805 ATCCAGCGCGCAGAGCGCTCCAGCAGACCGGTGATCGCGCGCGCGCGCGCGCGCGCTTC 1864
QY 329 -----ThrSerGlnLysAlaThrTyrSer 336
Db 1865 GCGCTCCAGCAGTGGCGCGCACACCGACGAGCGCGTCAACAGCAGCTGACCTACCGC 1924
QY 337 PheThrAlaThrAlaGlyLysProLeuLysIleSerLeuValTyrSerAsp----- 353
Db 1925 AACCTCGGCACCGACGACGTCAGCTGAGCTGACGTGACCGCGCGCGCGCGCGCGCGCG 1984
QY 354 -----AlaProAlaSer 357
Db 1985 AAGGCGCGCGCGCGCGCTTCTTACGCTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2044
QY 358 ThrThrAlaSerValThrLeuValAsnAspLeuAsnLeu----- 370
Db 2045 GGCAGCGCTCGCTGACATGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 2104
QY 371 -----ValIleThrAlaProAsnGlyThrGlnTyrVal 381
Db 2105 TACTCGCGCTACGTGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTC 2146

```

RESULT 14

US-09-000-016-1

; Sequence 1, Application US/09000016

; Patent No. 6143541

; GENERAL INFORMATION:

```

APPLICANT: AKIRA ARISAWA et al.
TITLE OF INVENTION: GENE ENCODING A PROTEIN HAVING ASYMMETRIC
TITLE OF INVENTION: HYDROLASE ACTIVITY FOR 4-SUBSTITUTED 1,4-DIHYDROPYRIDINE DERI
TITLE OF INVENTION: ITS EXPRESSION PRODUCT
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.
STREET: 2033 K Street, N.W., #800
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20006
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
COMPUTER: IBM Compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: Wordperfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/000,016
FILING DATE: January 30, 1998
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Warren M. Cheek, Jr.
REGISTRATION NUMBER: 33,367
REFERENCE/DOCKET NUMBER:
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-721-8200
TELEFAX: 202-721-8250
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2809 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: genomic DNA
ORIGINAL SOURCE:
ORGANISM: Streptomyces viridosporus
STRAIN: A-914
ORIGINAL SOURCE:
ORGANISM: Streptomyces antibioticus
STRAIN:
FEATURE:
NAME/KEY: CDS
LOCATION: 338...2539
IDENTIFICATION METHOD: E
FEATURE:
NAME/KEY: CDS
LOCATION: 2540...2809
IDENTIFICATION METHOD: P
US-09-000-016-1
Alignment Scores:
Pred. No.: 3,56e-26 Length: 2809
Score: 348.00 Matches: 129
Percent Similarity: 42.00% Conservative: 52
Best Local Similarity: 29.93% Mismatches: 167
Query Match: 15.49% Indels: 84
DB: 3 Gaps: 13
US-09-985-689A-1-COPY (1-434) x US-09-000-016-1 (1-2809)
QY 2 AspValAlaArgGlyIleValLysAlaAspValAlaGlnSerSerTyrGlyLeuTyrGly 21
Db 953 GACACGCTCGCGCGAGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1009
QY 22 GlnGlyGlnIleValAlaAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 41
Db 1010 AAGGCGGTGAAGATCGCGCGCTCTGGACACCGCGGTGTCACACGAGC----- 1054

```

QY 42 MetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsnAsn 61
Db 1055 ---CATCCGACCTGAAGCGCCGGGTGACCGGTCACAGAACTTCACCCCGCCGCGCCG 1111
QY 62 AlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGlySer 81
Db 1112 GCCGGCGACAGGTGGGCGACGCGACCCACGCTCCATCGCGCGCGGCGCGCGCC 1171
QY 82 ThrAsn-----LysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIle 97
Db 1172 CAGTCCAAAGGGAAGTACAGAGCGGTGCGACCGCGCGCGCGATCTCAACGCGCAAGTTC 1231
QY 98 MetAspSerGlyGlyLeuGlyLeuProSerAsnLeuGlnThrLeuPheSerGln 117
Db 1232 CTGACGACATCCCGT-----TTCCGGGACGACTCCGGCATCTCCCGCGCATGGAGTGG 1285
QY 118 AlaTyrSerAlaGlyAlaArgIle-HisThrAsnSerTrpGlyAlaAlaValAsnGlyAl 137
Db 1286 GCGGCGCGCAGGCGCGCGCGCGTCAACATGAGCTGGCGCGCATGGACACACCGGAG 1345
QY 137 aTyrThrAspSerArgAsnValAspTyrValArgLysAsnAspMetThrIleLe 157
Db 1346 ACCGACCCGCTGGAGGCGCG-GTGACAAAGCTGTCCGCGCGAAGGGCGTCTGTTCGC 1404
QY 157 upheAlaAlaGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLy 177
Db 1405 CATCGCGCGCGCAACGAGGCGCGGAG-----TCGATCGTTCGCGCGCGCGCGGA 1458
QY 177 sAsnAlaIleThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAs 197
Db 1459 CGCGCGCCCTCACCGTCGCGCGC-----GTGCA 1485
QY 197 pAsnIleAsnHisValAlaGlnPheSerArgGlyPro---ThrIlyAspGlyArgI 216
Db 1486 CGACAAGGACAGCTCGCGGACTCTCTCCACCGGCGCGCGCTCGCGCGCGCGCAT 1545
QY 216 eLysProAspValMetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaPr 236
Db 1546 CAAGCGGACGTCACCGCTCCCGCGGTGGACATCACCGCGCGCTCGCGCGGAGGCAACGA 1605
QY 236 aAspSerSerPheTrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMe 256
Db 1606 CATCGCGCAGAGGTGCGTGGAGGACCGCGCGGTATACATCATCTCCGCGCGCTCGAT 1665
QY 256 tAlaThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPheValIlyAsnAr 276
Db 1666 GCGCACCCCGCACGTCGCGCGCGCGCGCGCGCTCTGAAGCAGCAG----- 1711
QY 276 gGlyIleThrProLysProSerLeuLeuLysAlaAlaLeu-----IleAlaGlyAl 293
Db 1712 -----CACCCCGACTGACCTCCCGCGAAGTGAAGGCGCGCTCACCGGCTC 1758
QY 293 aAlaAspIleGlyLeuGlyTyrProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAs 313
Db 1759 CACCAAG---GGCGCAAGTACACCCGTTTCGAGCAGGGTTCGGCGCGGATCCAGGCGCA 1815
QY 313 pLysSerLeuAsnValAlaTyrValAsnGluSerSerSerLeuSer----- 328
Db 1816 CAAGCGCTCCAGCAGACCGTGTATCGCGACCGCGCTCTCGGTGAGCTTCGCGGTCCAGCA 1875
QY 329 -----ThrSerGlnLysAlaThrTyrSerPheThrAlaTh 340
Db 1876 GTGGCGCGCACACCGACGAGCGCGTCAACCAAGCAGCTGACTACCGCAACCTCGGCAC 1935
QY 340 tAlaGlyLysProLysLysIleSerLeuValTrpSerAsp----- 353
Db 1936 CCAGGACGTCAGCTGAAGCTGACGTACCGACCGCGCACCGACCCCAAGGCGGCGGCC 1995
QY 354 -----AlaProAlaSerThrThrAlaLase 361
Db 1996 GCGCGGCTTCTTCAGCTGGCGCGCACACCGGTGACCGTCCCGGCGGCGCGCGCTC 2055
QY 361 rValThrLeuValAsnAspLeuAsnLeu-----Va 371

Db 2056 CGTCGACATACCCCGACACCCGGCTCGCGCGCAGGTGAGCGCGTACTCGGCGTA 2115
QY 371 lileThrAlaProAsnGlyThrGlnTyrVal 381
Db 2116 CGTGGTCGCCACGCGCGCGCGCGAGACGTC 2146
RESULT 15
US-09-514-340-1
; Sequence 1, Application US/09514340
; Patent No. 6361987
; GENERAL INFORMATION:
; APPLICANT: Akira ARISAWA et al.
; TITLE OF INVENTION: GENE ENCODING A PROTEIN HAVING ASYMMETRIC
; HYDROLASE ACTIVITY FOR 4-SUBSTITUTED 1,4-DIHYDROPYRIDINE DE
; ITS EXPRESSION PRODUCT
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.
; STREET: 2033 K Street, N.W., #800
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20006
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/514,340
; FILING DATE: 28-Feb-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/000,016
; FILING DATE: January 30, 1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Lee Cheng
; REGISTRATION NUMBER: 40,949
; REFERENCE/DOCKET NUMBER: <Unknown>
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-721-8200
; TELEFAX: 202-721-8250
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2809 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
; ORIGINAL SOURCE:
; ORGANISM: Streptomyces antibioticus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 338...2539
; IDENTIFICATION METHOD: E
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 2540...2809
; IDENTIFICATION METHOD: P
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-514-340-1
Alignment Scores:
Pred. No.: 3,56e-26 Length: 2809
Score: 348.00 Matches: 129
Percent Similarity: 42.00% Conservative: 52
Best Local Similarity: 29.93% Mismatches: 167
Query Match: 15.49% Indels: 84
DB: 4 Gaps: 13

US-09-985-689A-1-COPY (1-434) x US-09-514-340-1 (1-2809)

QY 2 AspValAlaArgGlyIleValValAlaAspValAlaGlnSerSerTyrGlyLeuTyrGly 21
Db 953 GACAGTCCGTCGGCAGATCGGCGCCCAAGCGTGGTCGGC---GGCTACGACGGC 1009
QY 22 GlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSer 41
Db 1010 AAGGCGTGAAGATCGCGCTCTGACACACCGGTGTGACACGAGC----- 1054
QY 42 MethHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsnAsn 61
Db 1055 ---CATCGGACCTGAAGGCGGGGTGACCGGTCCCAAGACTTCACCGCGCCCGCCGC 1111
QY 62 AlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGlySer 81
Db 1112 GCCGGCGACAAAGTGGCGCACCGCACCCACCGCTCGCTCGATCGCGCGGCGCACGGCGCC 1171
QY 82 ThrAsn-----LysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIle 97
Db 1172 CAGTCCAAAGGCAAGTACAAAGGCGTGGCACCCCGCGCGGTGATCCTCAACGGCAAGGTC 1231
QY 98 MetAspSerGlyGlyLeuGlyLeuProSerAsnLeuGlnThrLeuPheSerGln 117
Db 1232 CTCGACGACTCCGGT-----TTGCGCAGCACTCCGCGATCCTCGCGCGATCGAGTGG 1285
QY 118 AlaTyrSerAlaGlyAlaArgIle-HisThrAsnSerTrpGlyAlaAlaValAsnGlyAl 137
Db 1286 GCGGCGCGCAGCGGCGCGACGTGTCACATGACCTGGCGGCGATGGACACACCGGAG 1345
QY 137 aTyrThrThrAspSerArgAsnValAspAspTyrValArgHisAsnAspMetThrIleLe 157
Db 1345 ACCGACCCGCTGGAGCGCGC-GTCGACAAAGTGTCCGCGCAGAGGCGTCTCTGTCGC 1404
QY 157 uPheAlaAlaGlyAsnGlyGlyProAsnGlyThrIleSerAlaProGlyThrAlaAl 177
Db 1405 CATCGCGCGCGCAACGAGGGCGCGAG-----TCGATCGTTCGCCGCGAGCGCGCA 1458
QY 177 sAsnAlaIleThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAs 197
Db 1459 CGCGCGCTCACGTCGCGCGC-----GTGCA 1485
QY 197 pAsnIleAsnHisValAlaGlnPheSerSerArgGlyPro---ThrLysAspGlyArgI 216
Db 1486 CGACAAGGACACAGCTCGCGACTTCTCTCCACCGCGCCCGCTCGCGGAGCGGCAACGA 1545
QY 216 eLysProAspValMetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaPr 236
Db 1546 CAAGCCGAGCTCACGCTCCGCGCTGGACATCACGCGCGCTCGCGGAGCGGCAACGA 1605
QY 236 oAspSerSerPheTrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyThrSerMe 256
Db 1606 CATCGGCGAGGAGTGGTGGAGGACCGCGCGCTACATGACCATCTCCGCGACGTCGAT 1665
QY 256 tAlaThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPheValIysAsnAr 276
Db 1666 GCGGACCCCGACGTCGCGGCGCGCGCGCTCTCTGAGCAGCAG----- 1711
QY 276 gGlyIleThrProLysProSerLeuLeuIysAlaAlaLeu-----IleAlaGlyAl 293
Db 1712 -----CACCCGACTGGACCTCCGCGCAACTGAAGGCGCGCTCACCGGCTC 1758
QY 293 aAlaAspIleGlyLeuGlyTyrProAsnGlyAsnGlnGlyTrpGlyArgValThrLeuAs 313
Db 1759 CACCAAG---GGCGGCAAGTACACCCCGTTCCGAGCAGGGTTCGGCGCGATCCAGGCCGA 1815
QY 313 pLysSerLeuAsnValAlaTyrValAsnGlnUserSerSerLeuser----- 328
Db 1816 CAAGCGCTCCAGCAGACCGGTGTCGCGCGACCCCGGTCTCGGTGAGCTTCGGCGTCCAGCA 1875
QY 329 -----ThrSerGlnLysAlaThrTyrSerPheThrAlaTh 340
Db 1876 GTGCGCGCACACCGACGCGGTGCACCAAGCAGCTGACCTACCGCAACCTCGGCAC 1935

340 rAlaGlyLysProLeuLysIleSerLeuValTrpSerAsp----- 353
1936 CCAGGACGTCACGCTGAAGCTGACGTGACCGCCACCGACCCCAAGGCAAGCGGCGCC 1995
354 -----AlaProAlaSerThrThrAla 361
1996 GCGGGCTTCCTCACGCTGGGCGCCACACGAGTGCACCTCCGCGCGGCGGCGGCGCTC 2055
361 rValThrLeuValAsnAspLeuAsnLeu-----Va 371
2056 CGTCGACATGACCGCGCGACACCCCGCTCGGCGCGGCGGCGGCGGCGCTC 2115
371 lIleThrAlaProAsnGlyThrGlnTyrVal 381
2116 CGTGTGCGCACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCTACTCGGCGTA 2115

Search completed: March 15, 2004, 22:29:59
Job time : 114 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: March 15, 2004, 21:48:38 ; Search time 360 Seconds
(without alignments)
4438.371 Million cell updates/sec

Title: US-09-985-689a-1-COPY

Perfect score: 2247

Sequence: 1 NDVARGIVKADVAQSSGLY.....EVQAVNPVGPQTRSLAIVN 434

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2432557 seqs, 184079884 residues

Total number of hits satisfying chosen parameters: 4865114

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlh
-Q=/cgn2_1/USFIO.spool/SWORE985689/runat_10032004_112808_19481/app_query.fasta_1.583
-DB=Published Applications NA -QWFI=fascap -SUFFIX=nrnp -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bts -START=1 -END=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR SCORE=ptc -THR MAX=100
-THR MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NOR=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USER=SWOP985689 @CGN 1.1.164 @runat_10032004_112808_19481
-NCPU=6 -ICPU=3 -NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -ICPU=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOF=6 -FGAPEXT=7 -YGAPOF=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications NA:

| | |
|-----|---|
| 1: | /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq: |
| 2: | /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq: |
| 3: | /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq: |
| 4: | /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq: |
| 5: | /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq: |
| 6: | /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq: |
| 7: | /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq: |
| 8: | /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq: |
| 9: | /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq: |
| 10: | /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq: |
| 11: | /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq: |
| 12: | /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq: |
| 13: | /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq: |
| 14: | /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq: |
| 15: | /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq: |
| 16: | /cgn2_6/ptodata/1/pubpna/US10D_PUBCOMB.seq: |
| 17: | /cgn2_6/ptodata/1/pubpna/US10E_PUBCOMB.seq: |
| 18: | /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq: |
| 19: | /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq: |

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | ID | Description |
|------------|-------|-------------|--------|----|-------------|
| ----- | | | | | |

| | | | | | | |
|----|-------|------|---------|----|----------------------|-------------------|
| 1 | 2242 | 99.8 | 1305 | 15 | US-10-385-662-1 | Sequence 1, Appli |
| 2 | 447.5 | 19.9 | 1977 | 13 | US-10-090-624-11 | Sequence 11, Appl |
| 3 | 411.5 | 18.3 | 1236 | 13 | US-10-090-624-2 | Sequence 2, Appli |
| 4 | 411.5 | 18.3 | 1236 | 13 | US-10-090-624-15 | Sequence 15, Appl |
| 5 | 363.5 | 16.2 | 3624 | 14 | US-10-156-761-5701 | Sequence 5701, Ap |
| 6 | 363.5 | 16.2 | 9025608 | 14 | US-10-156-761-1 | Sequence 1, Appli |
| 7 | 348 | 15.5 | 3417 | 14 | US-10-156-761-3306 | Sequence 3306, Ap |
| 8 | 348 | 15.5 | 9025608 | 14 | US-10-156-761-1 | Sequence 1, Appli |
| 9 | 336 | 15.0 | 135638 | 14 | US-10-314-657-1 | Sequence 1334, Ap |
| 10 | 332.5 | 14.8 | 1329 | 9 | US-09-974-300-1934 | Sequence 5, Appli |
| 11 | 307 | 13.7 | 4765 | 13 | US-10-090-624-5 | Sequence 5, Appli |
| 12 | 306.5 | 13.6 | 1560 | 15 | US-10-084-846A-113 | Sequence 113, App |
| 13 | 306.5 | 13.6 | 59816 | 15 | US-10-084-846A-2 | Sequence 2, Appli |
| 14 | 306.5 | 13.6 | 59816 | 15 | US-10-084-846A-2 | Sequence 2, Appli |
| 15 | 282 | 12.6 | 3788 | 10 | US-09-927-827-33 | Sequence 33, Appl |
| 16 | 278 | 12.4 | 3743 | 10 | US-09-927-827-29 | Sequence 29, Appl |
| 17 | 276 | 12.3 | 1306 | 9 | US-09-966-921A-1 | Sequence 1, Appli |
| 18 | 276 | 12.3 | 1330 | 9 | US-09-966-921A-5 | Sequence 5, Appli |
| 19 | 275 | 12.2 | 3303 | 14 | US-10-156-761-5384 | Sequence 5384, Ap |
| 20 | 268 | 11.9 | 2192 | 12 | US-10-424-599-112429 | Sequence 112429, |
| 21 | 256 | 11.4 | 1485 | 9 | US-09-974-300-1938 | Sequence 1938, Ap |
| 22 | 253 | 11.3 | 1971 | 9 | US-09-974-300-1935 | Sequence 1935, Ap |
| 23 | 251 | 11.2 | 840 | 14 | US-10-209-812-1 | Sequence 1, Appli |
| 24 | 247.5 | 11.0 | 2166 | 12 | US-10-344-231-17 | Sequence 17, Appl |
| 25 | 247.5 | 11.0 | 2166 | 12 | US-10-363-332A-17 | Sequence 17, Appl |
| 26 | 242.5 | 10.8 | 3884 | 10 | US-09-927-827-34 | Sequence 34, Appl |
| 27 | 240 | 10.7 | 1140 | 8 | US-08-322-678-11 | Sequence 11, Appl |
| 28 | 240 | 10.7 | 1140 | 16 | US-10-323-324-11 | Sequence 11, Appl |
| 29 | 240 | 10.7 | 1140 | 16 | US-10-323-324-12 | Sequence 12, Appl |
| 30 | 240 | 10.7 | 1143 | 14 | US-10-313-853-6 | Sequence 6, Appli |
| 31 | 240 | 10.7 | 2588 | 12 | US-10-344-231-20 | Sequence 20, Appl |
| 32 | 240 | 10.7 | 2588 | 12 | US-10-363-332A-20 | Sequence 20, Appl |
| 33 | 240 | 10.7 | 1497 | 8 | US-08-322-678-6 | Sequence 6, Appli |
| 34 | 237.5 | 10.6 | 1497 | 9 | US-09-060-854B-1 | Sequence 1, Appli |
| 35 | 237.5 | 10.6 | 1497 | 14 | US-10-033-325-1 | Sequence 1, Appli |
| 36 | 237.5 | 10.6 | 1497 | 14 | US-10-228-572-1 | Sequence 1, Appli |
| 37 | 237.5 | 10.6 | 1497 | 15 | US-10-423-649-1 | Sequence 1, Appli |
| 38 | 237.5 | 10.6 | 1497 | 16 | US-10-323-324-6 | Sequence 6, Appli |
| 39 | 237.5 | 10.6 | 1149 | 15 | US-10-146-905A-7 | Sequence 7, Appli |
| 40 | 235.5 | 10.5 | 1494 | 14 | US-10-104-693-1 | Sequence 1, Appli |
| 41 | 235.5 | 10.5 | 3452 | 10 | US-09-927-827-30 | Sequence 30, Appl |
| 42 | 233 | 10.4 | 1332 | 14 | US-10-156-761-5689 | Sequence 5689, Ap |
| 43 | 231.5 | 10.3 | 1140 | 15 | US-10-146-905A-9 | Sequence 9, Appli |
| 44 | 230.5 | 10.3 | 1140 | 9 | US-09-920-118-13 | Sequence 13, Appl |
| 45 | 227.5 | 10.1 | 1140 | 9 | US-09-920-118-13 | Sequence 13, Appl |

ALIGNMENTS

RESULT 1
US-10-385-662-1
; Sequence 1, Application US/10385662
; Publication No. US20040000432A1
; GENERAL INFORMATION:
; APPLICANT: OKUDA, MITSUYOSHI
; APPLICANT: SATO, TSUYOSHI
; APPLICANT: SAITO, KAZUHIRO
; APPLICANT: SUMITOMO, NOBUYUKI
; APPLICANT: IZAWA, YOSHIFUMI
; APPLICANT: SAKI, KATSUHIKA
; APPLICANT: KOBAYASHI, TOHRU
; APPLICANT: NOMURA, MASAFUMI
; TITLE OF INVENTION: Alkaline protease
; FILE REFERENCE: 234938USO
; CURRENT APPLICATION NUMBER: US/10385,662
; CURRENT FILING DATE: 2003-03-12
; PRIOR APPLICATION NUMBER: JP 2002-081428
; PRIOR FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: JP 2002-165987
; PRIOR FILING DATE: 2002-06-06
; PRIOR APPLICATION NUMBER: JP 2002-304230
; PRIOR FILING DATE: 2002-10-18
; PRIOR APPLICATION NUMBER: JP 2002-304231

```

; PRIOR FILING DATE: 2002-10-18
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1305
; TYPE: DNA
; ORGANISM: Bacillus sp. KSM-KP43
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1305)
; OTHER INFORMATION:
US-10-385-662-1

Alignment Scores:
Pred. No.: 5,69e-241 Length: 1305
Score: 2242.00 Matches: 433
Percent Similarity: 100.00% Conservative: 1
Best Local Similarity: 99.77% Mismatches: 0
Query Match: 99.78% Indels: 0
DB: 15 Gaps: 0

US-09-985-689A-1-COPY (1-434) x US-10-385-662-1 (1-1305)

QY 1 AsnAspValAlaArgGlyLeuValLysAlaAspValAlaGlnSerSerTyrGlyLeuTyr 20
Db 1 AATGATGTTGGCGTGGAAATGCTCAAGCGGATGGCTCAGAGCAGCTACGGGTGTAT 60
QY 21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyValAsnAspSer 40
Db 61 GGACAAGGACRATCGTAGCGGTTCGGCATACAGGCTGTGATACAGGTTCGCAATGACAGT 120
QY 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyValGthrAsn 60
Db 121 TCGATGCATGAAGCCTTCGCGGGGAAATTAATGCTATATATGCTTGGGACGACGAAT 180
QY 61 AsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGly 80
Db 181 AATGCCAATGATACGAATGCTCATGCTACGCGATGCTGGTGGCTCGTATTAGGAAACGCG 240
QY 81 SerThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSer 100
Db 241 TCCACTAATAAGGAATGGCGCTCAGGCGAATCTAGTCTTCCATCTATCATGGATAGC 300
QY 101 GlyGlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyrSer 120
Db 301 GGTGGGGGACTTGGAGGACTACCTTGAATCTGCAACCTTATTCAGGCAAGCATACAGT 360
QY 121 AlaGlyAlaArgIleHisThrAsnSerTyrGlyAlaAlaValAlaAsnGlyValAlaTyrThr 140
Db 361 GCTGGTCCAGAAATTCATCAAACTCTGGGAGCAGCAGTGAATGGGGCTTACACAACA 420
QY 141 AspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAla 160
Db 421 GATTCAGAAATGTGGATGACTATGTGCGCAAAATATGATGACGATCCTTTTCGCTGCC 480
QY 161 GlyAsnGlyLeuProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIle 180
Db 481 GGGAAATGAAGGACCGGACCGGAAACCATCAGTGACAGGACAGCAGCTTAAATCAATA 540
QY 181 ThrValGlyAlaThrGluAlaAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsn 200
Db 541 ACAGTCGGAGCTACGGAAACCTCCGCCCAAGCTTTGGGTCTTATCGGACAATATCAAC 600
QY 201 HisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysProAspVal 220
Db 601 CATGTGGCAGTCTCTTTCACGTGACCGGACCAAGGATGGGCGGATCAACCGGATGTC 660
QY 221 MetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerPhe 240
Db 661 ATGGACCGGAGACGTTCATCTATCATCAGCAGATCTTCTTCCACCGGATTCCTTC 720
QY 241 TrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIle 260

```

```

Db 721 TGGCGAACCATGACAGTAATAATGCATACATGGGTGGACGTCCTCATGGCTACACCGATC 780
QY 261 ValAlaGlyAsnValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrPro 280
Db 781 GTTGCTGGAACGTGGCACAGCTTCGTGAGCATTTTGTGAAAAACAGAGGATCACACA 840
QY 281 LysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAspIleGlyLeuGlyTyr 300
Db 841 AAGCCTTCTCTATTAAAGCGGCACTGATTCGCGTGCAGCTGACATCGGCTTGGCTAC 900
QY 301 ProAsnGlyAsnGlnGlyTyrGlyArgValThrLeuAspLysSerLeuAsnValAlaTyr 320
Db 901 CCGAACGGTAACCAAGGATGGGACAGTGAATTCGTAATAATCCCTGAACGTTGCCATAT 960
QY 321 ValAsnGluSerSerSerLeuSerThrSerGlnLysAlaThrTyrSerPheThrAlaThr 340
Db 961 GTGACGAGCTCAGTTCCTATCCACCGACCAAAAGCGAGCTACTCGTTTACTGCTACT 1020
QY 341 AlaGlyLysProLeuLysIleSerLeuValTyrSerAspAlaProAlaSerThrThrAla 360
Db 1021 GCGGCAAGCCTTGAATAATCTCCCTGGTATGCTGATGCCCTGCGAGCAACACTGCT 1080
QY 361 SerValThrLeuValAsnAspLeuAsnLeuValIleThrAlaProAsnGlyThrGlnTyr 380
Db 1081 TCGGTAAACGCTTGTCAATGATCTGGACCTTGCTATACCGCTCCAAATGGCACACAGTAT 1140
QY 381 ValGlyAsnAspPheThrSerProTyrAsnAspAsnTyrAspGlyArgAsnValGlu 400
Db 1141 GTAGGAAATGACTTTACTTCGCCATACATAATACTGGGATGGCGCAATAACGTAGAA 1200
QY 401 AsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGluValGlnAlaTyrAsn 420
Db 1201 AATGTATTATTAAATCACCACCAAGCGGAGCTATACAAATTGAGGTACAGGCTTATAAC 1260
QY 421 ValProValGlyProGlnThrPheSerLeuAlaIleValAsn 434
Db 1261 GTACCGGTTGGACACAGACCTTCTGTTGGCAATTGTGAAT 1302

RESULT 2
US-10-090-624-11
; Sequence 11, Application US/10090624
; Publication No. US20020132335A1
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: SHIMODO, Tomoko
; APPLICANT: ASADA, Kiyozo
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
; FILE REFERENCE: TAKAKURA=6
; CURRENT APPLICATION NUMBER: US/10/090,624
; CURRENT FILING DATE: 2002-03-06
; PRIOR APPLICATION NUMBER: 09/445,472
; PRIOR FILING DATE: 1999-12-06
; PRIOR APPLICATION NUMBER: 151969/1997
; PRIOR FILING DATE: 1997-06-10
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 11
; LENGTH: 1977
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
US-10-090-624-11

Alignment Scores:
Pred. No.: 1.4e-39 Length: 1977
Score: 447.50 Matches: 137
Percent Similarity: 44.66% Conservative: 68
Best Local Similarity: 29.85% Mismatches: 153
Query Match: 19.92% Indels: 101
DB: 18 Gaps: 18

```

```
US-09-985-689A-1-COPY (1-434) x US-10-090-624-11 (1-1977)
QY 8 VallyAalaaspValaaglnserserTyrGlyLeuTyrGlyGlnGlyGlnIleValala 27
Db 433 ATAGGGCGGATACCGTCTGGAACCTCCCTCGGTACGACGGAAGCGGTGTGGTGTGCC 492
QY 28 ValAalaaspThrGlyLeuAaspThrGlyArgAsnAaspSerSerMetHisGluAlaPheArg 47
Db 493 ATCGTCGATACGGGTATAGCGGGAAC-----CACCCCGATCTGAAG 534
QY 48 GlyLysIleThrAlaLeuTyr---AlaLeuGlyArgThrAsnAsnAlaAsnAaspThrAsn 66
Db 535 GGCAAGGTCTAGCGTGGTACGACGCGGTCAACGGCAGGTTCGACCCCTACGATGACCA 594
QY 67 GlyHisGlyThrHisValaIaGlySerValLeuGlyAsnGlySerThrAsnLys----- 84
Db 595 GGACAGGAACCCACGTTGCGGTATCGTTGCGGGAACCGGACGCTTACCTCCCACTAC 554
QY 85 ---GlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMet-----AapSerGly 101
Db 655 ATAGGGCTGCCCCCGCGGGAAGTCTGTCGGGTCAAGGTTCTCGGTCCGACGGTTGC 714
QY 102 GlyGlyLeuGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyrSerAla 121
Db 715 GGAAGCGTCCACCATCATCGCGGGTGTGACTGGGTCTCCAGAACAGGACAGTAC 774
QY 122 GlyAlaArgIle-----HisThrAsnSer 129
Db 775 GGGATAGGTCTATCAACCTCTCCCTCGGTCTCCCGAGAGTCCGACGGAACCGACTCC 834
QY 130 TrpGlyAlaValAalaAsnGlyAlaTyrThrThrAspSerArgAsnValAaspTyrVal 149
Db 835 CTCAGTCAGGCGGTCAACAACGCTGGGACGCC----- 867
QY 150 ArgLysAsnAaspMetThrIleLeuPheAlaIaGlyAsnGluGlyProAsnGlyGlyThr 169
Db 868 -----GGTATAGTCTGCGTCCGCGCGGCAACAGCGGCGGCAACACTACACC 918
QY 170 IleSerAlaProGlyThrAlaLysAsnAlaIleThrValGlyAlaThrGluAsnLeuArg 189
Db 919 GTCGGCTCACCGCGCGCGGAGCAAGGTATATAACCGTCGGTGCA----- 963
QY 190 ProSerPheGlySerTyrAlaAaspAsnIleAsnHisValAlaGlnPheSerArgGly 209
Db 964 -----GTTGACGACACGACACATCCGACGCTTCTCCACAGGGGA 1005
QY 210 ProThrLysAaspGlyArgIleLysProAaspValMetAlaProGlyThrPheIleLeuSer 229
Db 1006 CCGACCGGCGGACGAGGCTCAAGCGGAAGTCTGCGCCCGCGGTGACATCATAGCC 1065
QY 230 AlaArgSerSerLeuAlaProAaspSerSerPheTrpAlaAsnHisAaspSerLysTyrAla 249
Db 1066 CCGCGCGCCAGC-----GGAACCCAGCATGGGACCCCGGATAAACGACTACTACACC 1116
QY 250 TyrMetGlyGlyThrSerMetAlaThrProfileValaIaGly---AsnValAlaGlnLeu 268
Db 1117 AAGCGCTCTGGAACGACGATGCGCACCCGCGCACGTTTCGGGCGGTGGCGGCTCATCCTC 1176
QY 269 ArgGlnHisPheValLysAsnArgGlyIleThrProLys-----ProSerLeuLeuLys 286
Db 1177 CAGGCCCCAC-----CCGAGCTGGACCCCGGACAGGTGAAG 1212
QY 287 AlaAlaIleuIleAlaGlyAla-----AlaAaspIleGlyLeu 298
Db 1213 ACCGCCCTCATCGAGACCGCGCATAGTCCGCCCAAGGAGATAGCGGACATCGCCTAC 1272
QY 299 GlyTyrProAsnGlyAsnGlnGlyTyrGlyArgValThrLeuAaspLysSerLeu----- 316
Db 1273 GGTGG-----GGTAGGTGAACGTCTACAAGGGCCATCAAGTAC 1311
QY 317 ---AenValAlaTyrValAsnGluSerSerSerLeuSerThrSerGlnLysAlaThrTyr 335
Db 335
```

```
Db 1312 GACGACTACGCCAAGCTCACCTTACCGGCTCGGTCCGCGACAAAGGAGCGCCACCCAC 1371
QY 336 SerPheThrAlaThrAlaGlyLysProLeuLysIleSerLeuValTrpSerAspAlaPro 355
Db 1372 ACCTTCACGCTCAGCGCGCCACCTTCGTGACCGCACCTCTACTGGGAC----- 1422
QY 356 AlaSerThrThrAlaSerValThrLeuValAsnAaspLeuAsnLeuValIleThrAlaPro 375
Db 1423 -----ACGGGCTCGAGGACATCGACCTCTACCTCTACGACCCC 1461
QY 376 AsnGlyThrGlnTyrValGlyAsnAaspPheThrSerProTyrAsnAaspAsnTrpAspGly 395
Db 1462 AACGGGAACGAG-----GTTGACTACTCTCTACACCGCTACTAC----- 1500
QY 396 ArgAsnAsnValGluAsnValPheIleAsnAlaProGlnSerGlyThrTyrThrIleGlu 415
Db 1501 -----GCGTTCGAGAAGGTGCGTACTACAAACCGGACCGCGGAACTCGAGGTCGCAAG 1554
QY 416 ValGlnAlaTyrAsnValProValGlyProGlnThrPheSerLeuAlaIleValAsn 434
Db 1555 GTCGTCTAGCTACAAG-----GGCGGGCGGAACCTACGAGTCGACGTCGTCAGC 1602

RESULT 3
US-10-090-624-2
; Sequence 2, Application US/10090624
; Publication No. US20020132335A1
; GENERAL INFORMATION:
; APPLICANT: TAKAKURA, Hikaru
; APPLICANT: MORISHITA, Mio
; APPLICANT: SHIMOJO, Tomoko
; APPLICANT: ASADA, Kiyozo
; APPLICANT: KATO, Ikunoshin
; TITLE OF INVENTION: SYSTEM FOR EXPRESSING HYPERTHERMOSTABLE
; FILE REFERENCE: TAKAKURA-6
; CURRENT APPLICATION NUMBER: US/10/090,624
; CURRENT FILING DATE: 2002-03-06
; PRIOR APPLICATION NUMBER: 09/445,472
; PRIOR FILING DATE: 1999-12-06
; PRIOR APPLICATION NUMBER: 151969/1997
; PRIOR FILING DATE: 1997-06-10
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 1236
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
US-10-090-624-2

Alignment Scores:
Pred. No.: 5,81e-36 Length: 1236
Score: 411.50 Matches: 140
Percent Similarity: 42.55% Conservative: 60
Best Local Similarity: 25.75% Mismatches: 148
Query Match: 18.3% Indels: 122
DB: 13 Gaps: 20

US-09-985-689A-1-COPY (1-434) x US-10-090-624-2 (1-1236)
QY 12 ValAlaGlnSerSerTyrGly-LeuTyr-----GlyGlnGly 24
Db 24 GTCGCGAGCTCAAGTTATGGCACTTAGTTGGAACTGGGATATGATGTTCTGGAAT 83
QY 24 nIleValAlaValAalaAaspThrGlyLeuAaspThrGlyArgAsnAaspSerSerMetHisG 44
Db 84 CACAATAGGAATAATGACACTGGAATTGAC-----GCTTCTCATCC 125
QY 44 uAlaPheAGGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsnAsnAlaAsn-- 63
Db 126 AGATCTCCCAAGAAAGTA-----ATTGGTGGGTAGATTGTCATGCG 170
QY 64 -----AaspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuG 78
Db 78
```


[illegible]

```

; Sequence 5701, Application US/10156761
; Publication NO. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIEBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 5701
; LENGTH: 3624
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(3624)
US-10-156-761-5701

Alignment Scores:
Pred. No.: 7,02e-30 Length: 3624
Score: 363.50 Matches: 128
Percent Similarity: 45.50% Conservative: 54
Best Local Similarity: 32.00% Mismatches: 145
Query Match: 16.18% Indels: 73
DB: 14 Gaps: 18

US-09-985-689A-1-COPY (1-434) x US-10-156-761-5701 (1-3624)

Qy 8 VallyAlaAspValAlaGlnSerSer-----Tyr 17
      |||:|||||:|||||:|||||:|||||:
Db 556 GTCGAGGCGGCATGCGCGGAGACGACGCGAGATCGTACGCGGCGCGCGTGGGACGCC 615

Qy 18 GlyLeuTyGlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArg 37
      |||:|||||:|||||:|||||:|||||:|||||:
Db 616 GGGCTACGGGGGACGGCGTCACGTCGCGCGTGTCTACACCGCGGTCCACACC----- 669

Qy 38 AsnAspSerSerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrrAlaLeuGly 57
      |||:|||||:|||||:|||||:|||||:
Db 670 -----ACTCACCCGACCTCGCGCGCGGTGTCCCGGAGCAAGAGCTTCATC 717

Qy 58 ArgThrAsnAsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeu 77
      |||:|||||:|||||:|||||:|||||:
Db 718 GACGGGGAGAGTGTCGCGCACCGGACCGGACCCACGTCACCTCGACCGGTGCGC 777

Qy 78 GlyAsnGlySer-----ThrAsnLysGlyMetAlaProGlnAlaAsnLeuVal 93
      |||:|||||:|||||:|||||:
Db 778 GCGAGCGGCGCGCGCTCCGACGGACGCGGCGGTGCGCGCCGTCGCCACGCTCGCC 837

Qy 94 PheGlnSerIleMet---AspSerGlyGlyGlyLeuGlyGlyLeuProSerAsnLeuGln 112
      |||:|||||:|||||:|||||:
Db 838 GTCGGCAAAAGTGCTCAGCGCACCGGGGCGCGGA-----ACCGAGTCCCCAG 882

Qy 113 ThrLeuPheSerGlnAlaTy-SerAla-----GlyAlaArgIleHisThrAsnSer 129
      |||:|||||:|||||:|||||:
Db 883 ATCATCGCGGGCATGGAATGGCGCGCGGAGCGTGTCCGAGATGCTCTCGATGACG 942

Qy 130 TrpGlyAla-----AlaValAsnGlyAlaTyrrThrAsp-----SerArgAsnVal 145
      |||:|||||:|||||:|||||:
Db 943 CTCGGATCGACGAGGCCAGCGCGG-----ACCGACCCCATGGCGCGGCGCGTC 993

Qy 146 AspAspTyrrValArgLysAsnAspMetThrIleLeuPheAlaAlaGlyAsnGlyPro 165
      |||:|||||:|||||:|||||:
Db 994 GACACCCCTCTCGAGGACGACGGGCGCCCTCTTCGTGTCGCCGGGGAAACCGGTGCC 105

```

```

; OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1
Alignment Scores:          9 4e-25          Length:          9025608
Pred. No.:                36350           Matches:         128
Score:                    45.50%           Conservative:    54
Percent Similarity:      32.00%           Mismatches:    145
Best Local Similarity:   16.18%           Indels:        73
Query Match:             14              Gaps:          18
DB:
US-09-985-689A-1-COPY (1-434) x US-10-156-761-1 (1-9025608)

Qy      8  VallyAlaAspValAlaGlnSerSer-----Tyr 17
Db      6918913  GTGAGCGCGACATGCGGAGCAACCGCGAGATCGGTACGGGGCCCGGTGGGAGCGCC 6918972
Qy      18  GlyLeuTyrGlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyAArg 37
Db      6918973  GGCTCACGGGCGACGGGTACCGCTCGCGGTCTGCACACCGCGCTGCACACC----- 6918926
Qy      38  AsnAspSerSerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGly 57
Db      6918927  -----ACTCACCCCGACCTCGCGCGCGGTCTCCCGGAGCAAGAGCTTCATC 6918974
Qy      58  ArgThrAsnAsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaAlaGlySerValLeu 77
Db      6918975  GACGGGGAGGAGGTCCCGACCGCAACGCCACGGCCACCGTCACTCGACCGTCGCC 6919034
Qy      78  GlyAsnGlySer-----ThrAsnLysGlyMetAlaProGlnAlaAsnLeuVal 93
Db      6919035  GGCAGCGCGCCCGCTCCGACGGCAGCGGCGGTCTCGCGCGCGTCCCGCTCGCC 6919094
Qy      94  PheGlnSerIleMet---AspSerGlyGlyGlyLeuGlyGlyLeuProSerAsnLeuGln 112
Db      6919095  GTCGGAAAGTGTCTACGACACCGAGCGCGGGA-----AGCGAGTCCCGAG 6919139
Qy      113  ThrLeuPheSerGlnAlaTyrSerAla-----GlyAlaArgIleHisThrAsnSer 129
Db      6919140  ATCATCGCGGCGATGGAATGGCGCGCGGAGCGTGGTCCGAGGATCTCTCGATGAGC 6919199
Qy      130  TrpGlyAla-----AlaValAsnGlyAlaTyrThrAsp-----SerArgAsnVal 145
Db      6919200  CTGGATTCGACCGAGCGCGACCGGCGG-----ACGACCCCATGCGCGGCGCGTC 6919250
Qy      146  AspAspTyrValArgLysAsnAspMetThrIleLeuPheAlaAlaGlyAsnGluGlyPro 165
Db      6919251  GACACCCCTCTCCGAGGAGACCGCGCGCTCTTCGCTCGTCCGCGGGAACACCGGTGCC 6919310
Qy      166  AsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThrValGlyAlaThr 185
Db      6919311  ---CCCTCTCGATCGGCTCGCGCGCGCGCGCGCTCGCGTCCGCTGCGCGCGC--- 6919364
Qy      186  GluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsnHisValAlaGlnPhe 205
Db      6919365  -----GTGACTCATCCGACCGCGCGCTACTTC 6919394
Qy      206  SerSerArgGlyProThrLys---AspGlyArgIleLysProAspValMetAlaProGly 224
Db      6919395  ACCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 6919454
Qy      225  ThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheThrAlaAsnHis 244
Db      6919455  GTCGACATCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 6919502
Qy      245  AspSerLysTyrAlaTyrMetGlyThrSerMetAlaThrProIleValAlaGlyAsn 264
Db      6919503  -----TACACCTCCATGAGCGGTACGTGATGCGGACGCCCGCGCGCGCGCGTC 6919553
Qy      265  ValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrProLysProSerLeu 284
Db      6919554  GCCGCGCTCTCTCGCGGAGAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 6919613

```

```

Qy      166  AsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThrValGlyAlaThr 185
Db      1054  ---CCCTCTCGATCGGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1107
Qy      186  GluAsnLeuArgProSerPheGlySerTyrAlaAspAsnIleAsnHisValAlaGlnPhe 205
Db      1108  ---GTCGACTCATCCGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1137
Qy      206  SerSerArgGlyProThrLys---AspGlyArgIleLysProAspValMetAlaProGly 224
Db      1138  ACCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1197
Qy      225  ThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSerSerPheThrAlaAsnHis 244
Db      1198  GTCGACATCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1245
Qy      245  AspSerLysTyrAlaTyrMetGlyThrSerMetAlaThrProIleValAlaGlyAsn 264
Db      1246  ---TACACTCCATGAGCGGTACGTGATGCGGACGCCCGCGCGCGCGCGCGCGCG 1296
Qy      265  ValAlaGlnLeuArgGluHisPheValLysAsnArgGlyIleThrProLysProSerLeu 284
Db      1297  GCGCGCTCTCTCGCGGAGAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1356
Qy      285  LeuLysAlaAla-----LeuIleAlaGlyAlaAlaAspIleGlyLeuGlyTyrProAsn 302
Db      1357  ATGTCCACGTTCGAGCACTGACGCTCTCGATATATCATGTTGGGCGCGGTCTCGGT 1416
Qy      303  GlyAsnGlnGlyTrpGly---ArgValThrLeuAspLysSerLeuAsnValAlaTyr--- 320
Db      1417  GTCCGCGAGCGCGTGGCGCGCGCGTCAACCGGAGCGCGCGCGCGCGCGCGCGCG 1476
Qy      321  -----ValAsnGluSerSerSerLeuSerThrSerGln 331
Db      1477  CGTGCGCCCATGACCGCGATGACCGCTGACGAGACCGGTCACTTCACTTCACTTCC 1536
Qy      332  LysAlaThrTyrSerPheThr---AlaThrAlaGlyLysProLeuLysIleSerLeuVal 350
Db      1537  GACACGACGTGCGAGTTGAGCTCGCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCG 1596
Qy      351  TrpSerAspAla-----ProAlaSerThrThrAlaSerValThrLeuValAsnAsp 367
Db      1597  GCGGACACCGCACTACCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1656

```

RESULT 6

```

US-10-156-761-1
; Sequence 1, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
; LENGTH: 9025608
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (4187715)

```

```

QY 285 LeuLeuAlaAla-----LeuLeuAlaGlyAlaAlaAspIleGlyLeuGlyTyrProAsn 302
Db 6919614 ATGTCACCGTCCGAGCAACTCGACGCTCCGTATATACAGTTGGGGCGGTCGGGTCACT 6919673
QY 303 GlyAsnGlnGlyTrpGly---ArgValThrLeuAspLysSerLeuAsnValAlaTyr--- 320
Db 6919674 GTCCCGGACCGGTCGGCGCCGCGTCAACGCGGAGCGGCGCGGACCTCGGCTCCAC 6919733
QY 321 -----ValAsnGluSerSerSerLeuSerThrSerGln 331
Db 6919734 CGCTGCGCCCATGACGCGCATGACCGCTCACGAGACGCTCACTTCTCCAACTCTCTCC 6919793
QY 332 LysAlaThrTyrSerPheThr---AlaThrAlaGlyLysProLeuLysIleSerLeuVal 350
Db 6919794 GACACACGCTGCGAGTTGAGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 6919853
QY 351 TrpSerAspAla-----ProAlaSerThrThrAlaSerValThrLeuValAsnAsp 367
Db 6919854 GCGGACACCGCACTCAACCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 6919913

RESULT 7
US-10-156-761-3306
; Sequence 3306, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156, 761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 3306
; LENGTH: 3417
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(3417)
US-10-156-761-3306

Alignment Scores:
Pred. No.: 3,5e-28 Length: 3417
Score: 348.00 Matches: 139
Percent Similarity: 44.75% Conservative: 57
Best Local Similarity: 31.74% Mismatches: 170
Query Match: 15.49% Indels: 74
DB: 14 Gaps: 17

US-09-985-689A-1-COPY (1-434) x US-10-156-761-3306 (1-3417)
QY 18 GlyLeuTyrGlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArg 37
Db 745 GGGTACGACGCGCAGGCGGTCAAGATCGCGTCTGTGACACCGGTGTCGAC----- 795
QY 38 AsnAspSerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGly 57
Db 796 -----GGGACCCACCGGACCTCAAGGACCGAGGTGGCGCGGAGTCCCAAGAACTTCTCC 846
QY 58 ArgThrAsnAlaAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeu 77
Db 847 CCGCGCGCGGACGCGCGCGGACCTTGGTTCACGCGGACGACGTCGCGGTCCATCGCGGCG 906
QY 78 GlyAsnGlySerThrAsn-----LysGlyMetAlaProGlnAlaAsnLeuVal 93

```

```

Db 907 GGCACCGCCCAAGTCCAAAGCGCAAGTCAAGGGTGTCCGCGCGGCGGCGAGATCCTC 966
QY 94 PheGlnSerIleMetAspSerGlyGlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThr 113
Db 967 AACGGCAAGTCTCGACGACACCGGC-----TCCGCGACGACTCCGCGCATCTGGCC 1020
QY 114 LeuPheSerGlnAlaTyrSerAlaGlyAlaArgIle-HisThrAsnSerTrpGlyAlaAl 133
Db 1021 GGCATGAGTGGGGCGCGGAGCGGCGCGCGAGTGTCTCACTGAGCTGGGGGCGGCG 1080
QY 133 aValAsnGlyAlaTyrThrAspSerArgAsnValAspTyrValArgLysAsnAs 153
Db 1081 GACACCCCGAGATCGACCGCTGGAAGCGAG-GTCAACAAGCTCTCCGAGGAAGAGG 1139
QY 153 pMetThrIleLeuPheAlaGlyAsnGlyProAsnGly---GlyThrIleSerAl 172
Db 1140 CATCTCTTCGCGATCCGCGCGGCAACGAGGGAGTTCGGCGAGCGAGACCATCGGCTC 1199
QY 172 aProGlyThrAlaLysAsnAlaIleThrValGlyAlaThrGluAsnLeuArgProSerPh 192
Db 1200 CCGGGGACGCGCGCGCGCGCTCACCGTCGGCGC----- 1236
QY 192 eGlySerTyrAlaAspAsnIleAsnHisValAlaGlnPheSerSerArgGlyProThrLy 212
Db 1237 -----GTGAACGACGACGCGTCAACGAGTGTCTCTCCAGCGCGCGCGCGCGCT 1286
QY 212 sAspGlyArgIleLysProAspValMetAlaProGlyThrPheIleLeuSerAlaArgSe 232
Db 1287 GACGGCGCCATCAAGCCCGACGTCACCGACCGCGCGGTGGACATCACCGCGCGC----- 1341
QY 232 rSerLeuAlaProAspSerSerPheTrpAlaAsnHisAspSerLys-----TyrAl 249
Db 1342 ---GCGCGCGCGCGCGTCAACGAGTGTGGCGAGAGCGCGCGCGCGCGCGCGCTACCT 1397
QY 249 aTyrMetGlyGlyThrSerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeuAr 269
Db 1398 CACCATCTCCGAGTACGTGAGTGGCGACCCCGCATGTCCGCGCGCGCGCGCGATCCTCAA 1457
QY 269 gGluHisPheValLysAsnArgGlyIleThrProLysProSerLeu-----LeuLysAl 287
Db 1458 GCAGCAGCAC-----CCCAACTGGTGTCTCGCGCGAGCTCAAGGG 1496
QY 287 aAlaLeuIleAlaGlyAlaAlaAspIleGlyLeuGlyTyrProAsnGlyAsnGlnGlyTr 307
Db 1497 CGCGCTGACCGGCTCCGCGAAG-----GGCGGCAAGTACACGCGCTTCCAGCGAGGCTC 1550
QY 307 pGlyArgValThrLeuAspLysSerLeuAsnValAlaTyrValAsnGluSerSerLe 327
Db 1551 GGGCGGTATCGCGGTGCGAAGCGCATCAAGCAGTCCGTGATCGCAACCCCGCAACTCGGT 1610
QY 327 uSerThrSerGlnLysAlaThrTyrSerPhe-ThrAlaThrAlaGlyLysProLeuLysI 347
Db 1611 G-----AGCTTCGGATCCAGCAGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1646
QY 347 LeSerLeuValTrpSerAspAlaProAlaSerThrThrAlaSerValThrLeuValAsnA 367
Db 1647 CAAGCGGTTCACCGACGAGCTCACTACCGCAACTCGCGCAGCAGTGAGC-----TCAC 1700
QY 367 sPLeuAsnLeuVal-----IleThrAlaProAsnGlyThrG 379
Db 1701 G-CTGAACCTCGCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTTCT 1759
QY 379 InTyrValGly---AsnAspPheThrSerProTyrAsnAspAsnTrpAspGlyArgAsnA 398
Db 1760 TCAAGCTGGCGCGCAGAGGTACGCTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1807
QY 398 snValGlu-----AsnValPheIleAsnAlaProGlnSerGlyThrThrIleG 415
Db 1808 CGGTCCGATTCACGCTCAACAGAGCTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGGT 1867
QY 415 luValGlnAlaTyrAsnValProValGlyProGlnThrPheSerLeuAla 431

```



```
QY 148 TyrValArgLysAsnAspMetThrIleLeuPheAlaAlaGlyAsnGluClyProAsnGly 167
Db 1552 CTTACCGAAAGTACGGTGTGTATTTCGTAATAGTCGAGGAATGAAGTCTCGCAT 1611
QY 168 GlyThrIleSerAlaProGlyThrAlaLysAsnAlaIleThrValGlyAlaThrGluAsn 187
Db 1612 AACATCGTTGGAAGTCTGGTGTGCAACAAAGGCAATAACTGTGGAGCTCTGCA-- 1668
QY 188 LeuArgProSerPheGlySerTyrAlaAsp----- 197
Db 1669 GTGCCCAATTAAAGCTGGATTATGTTTCCCAAGCACTTGGATATCTGATTACTATGA 1728
QY 198 -----AsnIleAsnHisValAlaGlnPheSerSerArgGly 209
Db 1729 TTCTATTACTTCCCGCTACACAAGTIT---AGATAGCATCTTCTTCAAGCAGAGGG 1785
QY 210 ProThrLysAspGlyArgIleLysProAspValMetAlaProGlyThrPheIleLeuSer 229
Db 1786 CCGAGAAATAGATGTTGAAATAAAACCAATGTAGTGGCTCCAGGTTACGAAATTTACTCA 1845
QY 230 AlaArgSerSerLeuAlaProAspSerSerPheTrpAlaAsnHisAspSerLysTyrAla 249
Db 1846 TCCTGCCGATGGATGGCGAGTGACTTC----- 1878
QY 250 TyrMetGlyGlyThrSerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeuArg 269
Db 1879 ---ATGTCGGAACCTTCGATGGCTACTCCACATGTCAGCGGTGCTTGCACCTCTCAT 1935
QY 270 GluHisPheValLysAsnArgGlyIleThrProLysProSerLeuLeuLysAlaAlaLeu 289
Db 1936 AGCGGG---CAAAGGCCGAGGAATATACTACAACTCCAGATATAATTAAGAGTTCTT 1992
QY 290 IleAlaGlyAlaAlaAspIle-----GlyLeuGlyTyrProAsnGly 303
Db 1993 GAGAGCGGTCAACCTGGCTGAGGAGATCCATATACTGGGAGAGTACACTGAGCTT 2052
QY 304 AsnGlnGlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaTyrValAsnGlu 323
Db 2053 GACCAAGGTATGTTGTTAAGTTCACCAAGTCTCTGGGAAATC----- 2097
QY 324 SerSerSerLeuSerThrSerGlnLysAlaThrTyrSerPheThrAlaThrAlaGlyLys 343
Db 2098 -----CTTAAGGCTATAAAGCGCAC 2118
QY 344 ProLeuLysIleSerLeuValTrpSerAsp-AlaProAlaSerThrThrAlaSerValTh 363
Db 2119 ACTCTCCAAATTGTTGATCACTGGCGACAGACCTCTACAGCACTTTCGGAGT---AC 2175
QY 363 rLeuValAsnAspLeu-----AsnLeuValIleThrAlaProAs 376
Db 2176 TTGGTGTGGACGTTATAAGAGGTCTCTACGCAAGGACTCTATACCTGACATTTGCGAG 2235
QY 376 nGlyThr--GlnTyrValGlyAsn-----AspPheThrSerProTyrAsnAsp 391
Db 2236 TGGCACATTAAGTACGTAGGAGGACACGAGACGAGTACAGAACTTTTGAGATCTATGCA 2295
QY 392 AsnTrp-----AspGly-----ArgAsnAsnValGluAsnVal 402
Db 2296 CCATGGATTAGCCCTTTTTCAGTGGAAAGTGTAACTTCTAGAGAACATACCGAGTTTTC 2355
QY 403 Phe-----IleAsnAlaProGlnSerGlyThrTyr----- 412
Db 2356 CTTAGGTGGAATATGATGATGAGGGTCTTGAGCGAGTCTCTATGTTGGAAGGATAATC 2415
QY 413 -----ThrIleGluValGlnAlaTyrAsnValProValGlyPro 425
Db 2416 ATTGATGATCAACACCGCCAGTTATTGAACAGAGATCTTGAACACAAATTGTTATTTCCC 2475
QY 426 GlnThrPheSer 429
Db 2476 GAGAAGTTCACT 2487
```

RESULT 12

```
US-10-084-846A-113
; Sequence 113, Application US/10084846A
; Publication No. US2004006026A1
; GENERAL INFORMATION:
; APPLICANT: WEITNAUER, GABRIELE
; APPLICANT: MUHLENWEG, AGNES
; APPLICANT: TREPZER, AXEL
; APPLICANT: BECHTHOLD, ANDREAS
; TITLE OF INVENTION: AVILAMYCIN DERIVATIVES
; FILE REFERENCE: 1974-005
; CURRENT APPLICATION NUMBER: US/10/084, 846A
; CURRENT FILING DATE: 2003-02-25
; PRIOR APPLICATION NUMBER: PCT/EP01/09815
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: DE 101 09 166.4
; PRIOR FILING DATE: 2001-02-25
; NUMBER OF SEQ ID NOS: 120
; SOFTWARE: PatentIn ver. 3.2
; SEQ ID NO 113
; LENGTH: 1560
; TYPE: DNA
; ORGANISM: Streptomyces viridochromogenes
; FEATURE:
; OTHER INFORMATION: avix16 dna: partial sequence of coding strand 1; nucleotide 1
; OTHER INFORMATION: corresponds to nucleotide 56,198 of coding strand 1.
US-10-084-846A-113
```

Alignment Scores:

| | | | |
|------------------------|----------|---------------|------|
| Pred. No.: | 4,77e-24 | Length: | 1560 |
| Score: | 306.50 | Matches: | 116 |
| Percent Similarity: | 39.20% | Conservative: | 51 |
| Best Local Similarity: | 27.23% | Mismatches: | 154 |
| Query Match: | 13.64% | Indels: | 105 |
| DB: | 15 | Gaps: | 19 |

US-09-985-689A-1-COPY (1-434) x US-10-084-846A-113 (1-1560)

```
QY 21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db 451 GGGCAGGAGTGACGGCGTACGTATCGACACCGCGCTC-----CGC 492
QY 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyr---AlaLeuGlyArgThr 59
Db 493 ATCACCCACACGACGATTCGGCGCGCGGCTCTCTACGGTACGACGCGCATCGACACGAC 552
QY 60 AsnAsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsn 79
Db 553 AACACCGCCCGAGGACGCGCCACGCGCACGCGACGCGTGGCGCGCGCATCGCCGCGCAAC 612
QY 80 GlySerThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAsp 99
Db 613 GCC-----TACGGGGTCGCCAAGAAGCAAGATCGTAGGCGTCCGGCGTGTGAAC 663
QY 100 SerGlyGly-----GlyLeuGlyGlyLeuProSerAsn 110
Db 664 AACTCCGCGCAGGCGACACCGCCAGGTCTGTCGCCGCGCATCGACTGGTCCGCGCGAAC 723
QY 111 LeuGlnThrLeuPheSerGlnAlaTyr-SerAlaGlyAlaArgIleHisThrAsnSerTrp 130
Db 724 -----GCCGTCAAGCGCGCGTCCGCAACATGTCCTC 756
QY 131 GlyAlaAlaValAsnGlyAlaTyrThrThrAspSerArgAsnValAspAspTyrValArg 150
Db 757 GCGCGCGCGCGCACACGCGCCCTCGACACGCGCGCTAGCGAAC-----GCCATG 804
QY 151 LysAsnAspMetThrIleLeuPheAlaAlaGlyAsnGluGlyProAsnGlyGlyThrIle 170
Db 805 GCCTCCGCGCGTACCTTCGCGCGCGCGCGACGAGTGCAGCACCGCTCCACGAGG 864
QY 171 SerAlaProGlyThrAlaLysAsnAlaIleThrValGlyAlaThr-----GluAsnLeu 188
Db 865 TCA---CCCGCAGCGGTCAACGAGGCCATCAGGTTCGCGCGCGACGACCGAGCTCGAGCGCC 921
```



```
QY 189 ArgProSerPheGlySerTyrAlaAspAsnIleAsnHisValAlaGlnPheSerSerArg 208
Db 922 AAGCGCGGTACTCCAACTACGGCTCCGTCCTC----- 954
QY 209 GlyProThrLysAspGlyArgIleLysProAspValMetAlaProGlyThrPheIleLeu 228
Db 955 -----GACCTCTTCGCGCCCGGTTCTCCATACC 984
QY 229 SerAlaArgSerSerLeuAlaProAspSerPheTyrAlaAsnHisAspSerLysTyr 248
Db 985 TCGGCC-----TGGAACTCAAGGACTCGCGGACC 1014
QY 249 AlaTyrMetGlyThrSerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeu 268
Db 1015 AACACATCTCCGGTACGTGATGGACCCCGACGTGGCGGCGCGCGCTC--- 1071
QY 269 ArgGluHisPheValLysAsnArgGlyIleThrProLys-----ProSerLeuLeu 285
Db 1072 -----CACCTCGCGCAACCCCTCGGCCACCCGTCCTCCAGGTCCGACCGCGGTGACG 1125
QY 286 LysAlaLeuIleAlaGlyAlaAlaAspIleGlyLeuGlyTyrProAsnGlyAsnGln 305
Db 1126 TCGCGCGCACACCGCGGTCTGTCACCAACCCCGCGACGGGTCTCGCCCAAC----- 1176
QY 306 GlyTyrGlyArgValThrLeuAspLysSerLeuAsnValAlaTyrValAsnGlu----- 323
Db 1177 -----CGGCTCTCTGACTGTCGGCGCGGCGCAC 1203
QY 324 -----SerSerSerLeuSerThrSerGlnLysAlaThrTyrSerPheThrAla 339
Db 1204 GACCACCTCCGCGCGCGGTTCGAGAACCCGGTGACTACACGATCAGCGACAACTCC 1263
QY 340 ThrAlaGlyLysProLeuLysIleSerLeuValTyrSerAspAlaProAlaSerThrThr 359
Db 1264 ACGTCTGAGTCCCGGTGAGGTCTCGGCGTCTCGGCAACCGCGCTTCGCGCTCGCC 1323
QY 360 AlaSerVal-----ThrLeuValAsnAspLeuAsnValIleThrAlaPro 375
Db 1324 GTAGAGTGCACATCGTCACACGTATCGGCGACCTCCAGGTCCAGCTGATCGCCGCC 1383
QY 376 AsnGlyThrGlnTyrValGlyAsnAspPhe---ThrSerProTyrAsnAspAsnTrpAsp 394
Db 1384 GAGGACGCGGTACAGCTCAAGTCTGACGACCGCGCGGCGGAGTTCGGAC----- 1434
QY 395 GlyArgAsnAsnValGluAsnValPhe---IleAsnAlaProGln-----SerGly 410
Db 1435 -----AACATCAACACACGTACTCGGTGACGCTCTCTCGGAGCGCGCAACGCG 1485
QY 411 ThrTyrThrIleGluVal 416
Db 1486 ACGTGGAACTGCGGGTG 1503

RESULT 13
US-10-084-846A-1
; Sequence 1, Application US/10084846A
; Publication NO. US20040006026A1
; GENERAL INFORMATION:
; APPLICANT: WEITNAUER, GABRIELE
; APPLICANT: MUHLNENEG, AGNES
; APPLICANT: TREFFER, AXEL
; APPLICANT: BECHTHOLD, ANDREAS
; TITLE OF INVENTION: AVILAMYCIN DERIVATIVES
; FILE REFERENCE: 1974-005
; CURRENT APPLICATION NUMBER: US/10/084, 846A
; CURRENT FILING DATE: 2003-02-25
; PRIOR APPLICATION NUMBER: PCT/EP01/09815
; PRIOR FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: DE 101 09 166.4
; PRIOR FILING DATE: 2001-02-25
; NUMBER OF SEQ ID NOS: 120
; SOFTWARE: Patent in Ver. 3.2
; SEQ ID NO 1
; LENGTH: 59816
```

```
; TYPE: DNA
; ORGANISM: Streptomyces viridochromogenes
US-10-084-846A-1

Alignment Scores:
Pred. No.: 1,17e-21 Length: 59816
Score: 306.50 Matches: 116
Percent Similarity: 39.20% Conservative: 51
Best Local Similarity: 27.23% Mismatches: 154
Query Match: 13.64% Indels: 105
DB: 15 Gaps: 19

US-09-985-689A-1-COPY (1-434) x US-10-084-846A-1 (1-59816)
QY 21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db 56648 GGGCAGGGAGTACGGCGGTACGTATCGACACCGCGCTC-----CGC 56689
QY 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTyr---AlaLeuGlyArgThr 59
Db 56690 ATCACCACACAGGACITTCGGCGCGCGGCTCTACGGCTAGACGCCATCGACACGAC 56749
QY 60 AsnAsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsn 79
Db 56750 AACACCGCCAGACCGCCACCGCGCGGACCGACCGTGGCGCGGACGTGCGCGGCAAC 56809
QY 80 GlySerThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAsp 99
Db 56810 GCC-----TACGGGTGCGCCAAAGAGCCCAAGATCGTAGGCTCGCGGTCTGAAC 56860
QY 100 SerGlyGly-----GlyLeuGlyGlyLeuProSerAsn 110
Db 56861 AACTCCGCGCAGGACACCCGCCAGGTGTCGCGCGGTACGTGGGTTCGCGCGGCAAC 56920
QY 111 LeuGlnThrLeuPheSerGlnAlaTyrSerAlaGlyAlaArgIleHisThrAsnSerTrp 130
Db 56921 -----GCGGTCAAGCGCGCTCGCCCAACATGTCCCTC 56953
QY 131 GlyAlaValAlaGlyAlaTyrThrThrAspSerArgAsnValAspAspTyrValArg 150
Db 56954 GGCAGCGCGCGCACACCGCGCTCGACACCGCGCTACGCAAC-----GCCATG 57001
QY 151 LysAsnAspMetThrIleLeuPheAlaAlaGlyAsnGluGlyProAsnGlyGlyThrIle 170
Db 57002 GCCTTCGCGGTACCTTCGCGGTGCGCGCGGACGAGTCGACCAACGCTCCACGAGG 57061
QY 171 SerAlaProGlyThrAlaLysAsnAlaIleThrValGlyAlaThr-----GluAsnLeu 188
Db 57062 TCA---CCCGCACGCGTCACCGAGGCCATCAGGTTCGCGCGGACGACGCTCGGACGCC 57118
QY 189 ArgProSerPheGlySerTyrAlaAspAsnIleAsnHisValAlaGlnPheSerSerArg 208
Db 57119 AAGCGCGGTACTCCAACTACGGCTCGCTCCTC----- 57151
QY 209 GlyProThrLysAspGlyArgIleLysProAspValMetAlaProGlyThrPheIleLeu 228
Db 57152 -----GACCTCTTCGCGCCCGGTTCTCCATACC 57181
QY 229 SerAlaArgSerSerLeuAlaProAspSerSerPheTrpAlaAsnHisAspSerLysTyr 248
Db 57182 TCGGCC-----TGGAACTCAAGGACTCGCGGACC 57211
QY 249 AlaTyrMetGlyThrSerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeu 269
Db 57212 AACACATCTCCGGTACGTGATGGACCCCGACGTGGCGGCGCGCGCTC--- 57268
QY 269 ArgGluHisPheValLysAsnArgGlyIleThrProLys-----ProSerLeuLeu 285
Db 57269 -----CACCTCGCGCAACCCCTCGGCCACCCGTCCTCCAGGTTCGCGCGGCTGACG 57322
QY 286 LysAlaAlaLeuIleAlaGlyAlaAlaAspIleGlyLeuGlyTyrProAsnGlyAsnGln 305
Db 57323 TCGCGCGCACCGCGGTCTGTCACCAACCCCGCGCGGCTCGCGCAAC----- 57373
```

```

QY 306 GlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaIleThrValAsnGlu----- 323
Db 57374 -----CGGCTCTGTACGTGCGCGGGGCAC 57400
QY 324 -----SerSerSerLeuSerThrSerGlnLysAlaThrTySerPheThrAla 339
Db 57401 GACCACCCCTCCGGCGCGGCTTCGAGAACACCGGTGACTACACGATCAGCGACAACTCC 57460
QY 340 ThrAlaGlyLysProLeuLysIleSerLeuValTrpSerAspAlaProAlaSerThrThr 359
Db 57461 ACGTTCAGTCCCGGTGAGCGTCTCGGCTCTCCGGCAACGGCGCTCCGCTCCGCTCCG 57520
QY 360 AlaSerVal-----ThrLeuValAsnAspLeuValIleThrAlaPro 375
Db 57521 GTAGAGTCCACATCGTCCACAGTACATCGGCGACCTCCAGTCCAGCTGATCGCCCC 57580
QY 376 AsnGlyThrGlnTyValGlyAsnAspPhe-----ThrSerProTyAsnAspAsnTrpAsp 394
Db 57581 GACGGCAGCGGTACACGCTCAAGTCAAGTCCGCGGACCGCGGCGAGTTCGGAC----- 57631
QY 395 GlyArgAsnAsnValGluAsnValPhe-----IleAsnAlaProGln-----SerGly 410
Db 57632 -----AACATCAACACCGTACTCGGTGACGCTCTCTCGAGGCGGCGCAACGCG 57682
QY 411 ThrTyThrIleGluVal 416
Db 57683 ACGTGGAACTCGGGTG 57700

RESULT 14
US-10-084-846A-2/c
; Sequence 2, Application US/10084846A
; Publication No. US20040006026A1
; GENERAL INFORMATION:
; APPLICANT: WEITNAUER, GABRIELE
; APPLICANT: MOHLEWEG, AGNES
; APPLICANT: TREFFER, AXEL
; APPLICANT: BECHTHOLD, ANDREAS
; TITLE OF INVENTION: AVILAMYCIN DERIVATIVES
; FILE REFERENCE: 1974-005
; CURRENT APPLICATION NUMBER: US/10/084,846A
; CURRENT FILING DATE: 2003-02-25
; PRIOR FILING DATE: 2001-08-24
; PRIOR FILING DATE: 2001-02-25
; NUMBER OF SEQ ID NOS: 120
; SOFTWARE: PatentIn Ver. 3.2
; SEQ ID NO 2
; LENGTH: 59816
; TYPE: DNA
; ORGANISM: Streptomyces viridochromogenes
US-10-084-846A-2

Alignment Scores:
Pred. No.: 1,17e-21 Length: 59816
Score: 306.50 Matches: 116
Percent Similarity: 39.20% Conservative: 51
Best Local Similarity: 27.23% Mismatches: 154
Query Match: 13.64% Indels: 105
DB: 15 Gaps: 19

US-09-985-689A-1-COPY (1-434) x US-10-084-846A-2 (1-59816)
QY 21 GlyGlnGlyGlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSer 40
Db 3169 GGGCAGGAGTGACGGCGTACGTATCATGACACCGCGCTC-----CGC 3128
QY 41 SerMetHisGluAlaPheArgGlyLysIleThrAlaLeuTy-----AlaLeuGlyArgThr 59
Db 3127 ATCAACCCACGCGACTTCGGCGCGCGGCTTCCTACGCTACGACGCGCTCGACACGAC 3068
QY 60 AsnAsnAlaAsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsn 79

```

```

Db 3067 AACACCCCGCAGGAGCGCCACGCGCCACGCGACGCGCTGCGCGCGGCAAC 3008
QY 80 GlySerThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAsp 99
Db 3007 GCC-----TACGCGCTCGCCAGAGAGCCAGATCGTAGCGCTCCGCTGCTGAAC 2957
QY 100 SerGlyGly-----GlyLeuGlyGlyLeuProSerAsn 110
Db 2956 AACTCCCGCCAGGCGCACACCGCCAGGTGTCGCCGCGCATCGACTGGTGGTCCGCGGCAAC 2897
QY 111 LeuGlnThrLeuPheSerGlnAlaTySerAlaGlyAlaArgIleHisThrAsnSerTrp 130
Db 2896 -----GCCGTCAAGCGCGCGTCCGCAACATCGCTCCTC 2854
QY 131 GlyAlaAlaValAsnGlyAlaTyThrThrAspSerArgAsnValAspAspTyValArg 150
Db 2863 GCGCGCGCGCGCGACACGCGCCCTCGACACGCGCGCTACGCAAC-----GCCATG 2816
QY 151 LysAsnAspMetThrIleLeuPheAlaAlaGlyAsnGluGlyProAsnGlyGlyThrIle 170
Db 2815 CCCTCCGCGCTCACCTTCGCCGCTGCGCGCGCAACGAGTCGACCAACGCTCCACGAGG 2756
QY 171 SerAlaProGlyThrAlaLysAsnAlaIleThrValGlyAlaThr-----GluAsnLeu 188
Db 2755 TCA---CCCGCACGCGTCACCGAGGCGCATCACGCTCGCGCGCGACGACGCTCGGCGCC 2699
QY 189 ArgProSerPheGlySerTyAlaAspAsnIleAsnHisValAlaGlnPheSerSerArg 208
Db 2698 AAGCGCGCTACTCCAACTACGCTCGCTCCTC----- 2666
QY 209 GlyProThrLysAspGlyArgIleLysProAspValMetAlaProGlyThrPheIleLeu 228
Db 2665 -----GACCTCTTCGCCCGCGTTCGCTCCATCACC 2636
QY 229 SerAlaArgSerSerLeuAlaProAspSerSerPheTrpAlaAsnHisAspSerLysTy 248
Db 2635 TCGGCC-----TGGAACCAAGCGACTCGCGGACC 2606
QY 249 AlaTyMetGlyThrSerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeu 268
Db 2605 AACACATCTCCGGTACGTGATGCGACCGCGACCGCGCGCGCGCGCGCGCTC--- 2549
QY 269 ArgGluHisPheValLysAsnArgGlyIleThrProLys-----ProSerLeuLeu 285
Db 2548 -----CACCTCGCGCGCAACCTTCGGCGCACCCCGTCCCGAGTCGCCACGCGCTGACG 2495
QY 286 LysAlaAlaLeuIleAlaGlyAlaAlaAspIleGlyLeuGlyTyProAsnGlyAsnGln 305
Db 2494 TCGCGCGCACCGCGCGTCTGTCACCAACCGCGCGCGCGCTCGCCCAAC----- 2444
QY 306 GlyTrpGlyArgValThrLeuAspLysSerLeuAsnValAlaTyValAsnGlu----- 323
Db 2443 -----CGGCTCTGTACGTGCGCGCGGCGAC 2417
QY 324 -----SerSerSerLeuSerThrSerGlnLysAlaThrTySerPheThrAla 339
Db 2416 GACACCCCTCCGGCGCGCGCTTCGAGAACACCGGTGACTACACGATCAGCGACAACTCC 2357
QY 340 ThrAlaGlyLysProLeuLysIleSerLeuValTrpSerAspAlaProAlaSerThrThr 359
Db 2356 ACGGTGAGTCCCGGTGACGCTCTCCGGCGCTCTCCGGCAACGCGCGCTCGGCGCTCGCC 2297
QY 360 AlaSerVal-----ThrLeuValAsnAspLeuAsnLeuValIleThrAlaPro 375
Db 2296 GTAGAGTCCACATCGTCCACAGTACATCGCGGACCTCCAGGTCCAGCTGATCGCCCC 2237
QY 376 AsnGlyThrGlnTyValGlyAsnAspPhe-----ThrSerProTyAsnAspAsnTrpAsp 394
Db 2236 GACGGCAGCGGTACACGCTCAAGTCTGACGCGACCGCGCGCGAGTTCGGAC----- 2186
QY 395 GlyArgAsnAsnValGluAsnValPhe-----IleAsnAlaProGln-----SerGly 410

```

Db 2185 -----AACATCAACACACGACTGCTGTAACGCTCTCTCGGAGCGGCCAACCGC 2135
 Qy 411 ThrTyrThrIleGluVal 416
 Db 2134 ACGTGAACACGCGGGTG 2117
 RESULT 15
 US-09-927-827-33
 ; Sequence 33, Application US/09927827
 ; Publication No. US20030036176A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Bower, Stanley G.
 ; TITLE OF INVENTION: Directed Genetic Engineering of Xanthomonas campestris
 ; FILE REFERENCE: 38-10(15824)B
 ; CURRENT APPLICATION NUMBER: US/09/927,827
 ; CURRENT FILING DATE: 2001-08-10
 ; PRIOR APPLICATION NUMBER: US 60/279,493
 ; PRIOR FILING DATE: 2001-03-28
 ; NUMBER OF SEQ ID NOS: 69
 ; SEQ ID NO 33
 ; LENGTH: 3788
 ; TYPE: DNA
 ; ORGANISM: Xanthomonas campestris
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (1001)..(2788)
 US-09-927-827-33

Alignment Scores:
 Pred. No.: 1,01e-20 Length: 3788
 Score: 282.00 Matches: 112
 Percent Similarity: 40.86% Conservatives: 49
 Best Local Similarity: 28.43% Mismatches: 113
 Query Match: 12.55% Indels: 120
 DB: 10 Gaps: 19
 US-09-985-689a-1-COPY (1-434) x US-09-927-827-33 (1-3788)
 Qy 3 ValAlaArgGlyLeuValAlaAspValAlaGlnSerTyrGlyLeuTyrGlyGln 22
 Db 1739 GTCTCGGGCAACGTGTGTGATGTGGATGACGACGACGCGGTATCGCCAGTGGGAG 1798
 Qy 23 GlyGlnIleValAlaValAlaAspThrGlyArgAsnAspSerMet 42
 Db 1799 CAGATCGGCTGGGGTCTCTGATACCGGATT-----GGGCGCGG 1840
 Qy 43 HisGluAlaPhe-----ArgGlyIleThrAlaLeuTyr----- 54
 Db 1841 CATCCGCATTCTCTGTCGCGCGGACGACATGTGGTGGCGCAATGGGATTCACG 1900
 Qy 55 -----AlaLeuGlyArgThrAsnAsnAlaAsnAsp 64
 Db 1901 CGGGCGGGCGGCCAAGCGGTGACGCGGCGCGATGGCAACGGTTACCGCACTCGAC 1960
 Qy 65 ThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGly----- 78
 Db 1961 GGGCAGCGCCATGCGACGACATCGCGGCGCATCATCGCGGCTGTTCCGCGCGCGGTG 2020
 Qy 79 -----AsnGlySerThrAsnLys-----GlyMetAlaProGlnAla 90
 Db 2021 CCTGATCCAGTGGCAAGCGCGGAACTGCTGGAATTCGCGCATGGCGCGGAGACG 2080
 Qy 91 AsnLeuVal---PheGlnSerIleMetAspSerGlyGlyLeuGly----- 106
 Db 2081 CAATCTATGGCTTCAAGGTGCTGGACGACGCGCGCAACGCGGCGGATTCGTGGATGATC 2140
 Qy 107 ---LeuProSerAsnLeuGlnThrLeuPheSerGlnAla----- 118
 Db 2141 AAGCAGTGGACGACGCGGCGGATCTCAACAGCGGTGCGCGGAGCTGGTGTATCCAGCGC 2200
 Qy 119 -----TyrSerAlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaValAsnGly 136

Db 2201 GTCAATCTCAGCTGGGCGGGTACTTCGATCCGGAAGCTACGCTGTGGCTTC----- 2254
 Qy 137 AlaTyrThrThrAspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIle 156
 Db 2255 -----ACGCGGTATGCAACGAGTTGCGCGGTTATGGCGGCGGCGTACTGTA 2305
 Qy 157 LeuPheAlaAlaGlyAsnGluGly-----ProAsnGlyGlyThr--- 169
 Db 2306 GTGGTGGCGCGCGCAACGAAGGCTGGCTGGCTGATGTCAGAACGACGCGCGGACCTAT 2365
 Qy 170 -----IleSerAlaProGlyThrAlaLysAsnAlaIleThrVal 182
 Db 2366 CCGCGCAACATGATCTGTCGATCAGCGATCCGGCAATCTGGAGGACGCGATCGTGGTG 2425
 Qy 183 GlyAlaThrGluAsnLeuArgPro---SerPheGlySerTyrAlaAspAsnIleAsnHis 201
 Db 2426 GGATCGGTGCACAGACGACCGCGCAATACGGC----- 2461
 Qy 202 ValAlaGlnPheSerArgGlyProThrLysAspGlyArgIleLysProaspValMet 221
 Db 2462 GTGTCGTATTTTTCATCCCGCGCGGACCGCGGATGGCGCTCCAAACCTGATGTGTC 2521
 Qy 222 AlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProaspSerSerPheTrp 241
 Db 2522 GCCCGCGCGGAAAAAGATCTCTCGCTTACTACGGCTTCGACCCG----- 2566
 Qy 242 AlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIleVal 261
 Db 2567 CGGACCCGTCACGCTGATGTCGAGATGAGCGGCGCACCGCATGCCCGACCGCATGTG 2626
 Qy 262 AlaGly-----Asn 264
 Db 2627 TCGGCGTGTGGCGCGGTTTTTATCCGACGCGCGGAGTTATCCGCTTCCGACCGG 2686
 Qy 265 ValAlaGlnLeu-----ArgGluHisPheValLysAsn 275
 Db 2687 GTCAAGCACTGCTGTGGACACTGCACCGACCTGCAGCGCGCATGTTACGTGCGAGGC 2746
 Qy 276 ArgGlyIleThrProLysProSerLeuLeuLysAlaAlaLeuIleAlaGlyAlaAlaAsp 295
 Db 2747 AGGGGGGTG-----CGAATTTGATCGG-----ATGCTTGGAGAGACGTGA 2788
 Qy 296 IleGly-----LeuGlyTyrProAsnGlyAsnGlnGly 306
 Db 2789 TTCGGAGTCGGGATTTGGGATTCGCAACCGCGTGGTTGGC 2830

Search completed: March 16, 2004, 01:21:14
 Job time : 7202 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: March 15, 2004, 20:30:22 ; Search time 2385 Seconds
(without alignments)
5434.040 Million cell updates/sec

Title: US-09-985-689a-1-COPY

Perfect score: 2247

Sequence: 1 NDVARGIVKADVAQSSYGLY.....EVQAVNPVGPQTFLAIVN 434

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+_p2n.model -DBV=xlh
-Q=/cgm2_1/USPTO.spool/SWOPE985689/runat_10032004_112806_19374/app_query.fasta_1.583
-DB=EST -QMT=FASTAP -SUFFIX=est -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR=SCORE=0 -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=EXC -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=SWOPE985689@cgm2_1_2135@runat_10032004_112806_19374 -NCPU=6 -ICPU=3
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

EST:
1: em_estba:*
2: em_estin:*
3: em_estin:*
4: em_estin:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: gb_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1.*

29: gb_gss2.*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID | Description |
|------------|-------|-------------|--------|----|-----------|--------------------|
| 1 | 246 | 10.9 | 1605 | 13 | BQ622771 | CC Contig BQ622771 |
| 2 | 241 | 10.7 | 640 | 12 | BJ395336 | BJ395336 |
| 3 | 226.5 | 10.1 | 4188 | 11 | AK029048 | AK029048 Mus muscu |
| 4 | 224.5 | 10.0 | 594 | 12 | BJ393752 | BJ393752 |
| 5 | 218.5 | 9.7 | 532 | 29 | TA319G10P | TA319G10P |
| 6 | 207.5 | 9.2 | 2141 | 13 | BQ142519 | BQ142519 T. brucei |
| 7 | 200 | 8.9 | 508 | 28 | AQ52212 | AQ52212 Sheared M |
| 8 | 199.5 | 8.9 | 601 | 12 | BJ387574 | BJ387574 |
| 9 | 195.5 | 8.7 | 1002 | 29 | CNS0606B | AL393417 T3 end of |
| 10 | 195 | 8.7 | 771 | 14 | CA320325 | CA320325 UI-M-FW0- |
| 11 | 191 | 8.5 | 633 | 12 | BJ369190 | BJ369190 |
| 12 | 190 | 8.5 | 718 | 12 | BI750157 | BI750157 |
| 13 | 186.5 | 8.3 | 4662 | 11 | BC060627 | BC060627 Mus muscu |
| 14 | 185.5 | 8.3 | 574 | 29 | TA315H10P | AL490202 T. brucei |
| 15 | 182.5 | 8.1 | 614 | 9 | AJ273402 | AJ273402 |
| 16 | 181 | 8.1 | 716 | 28 | BZ893395 | BZ893395 HL2 0177 |
| 17 | 180 | 8.0 | 641 | 12 | BJ393925 | BJ393925 |
| 18 | 180 | 8.0 | 665 | 13 | BQ770462 | BQ770462 UI-M-P10- |
| 19 | 179.5 | 8.0 | 3091 | 11 | BC011275 | BC011275 Mus muscu |
| 20 | 177.5 | 7.9 | 650 | 9 | AJ274038 | AJ274038 |
| 21 | 177.5 | 7.9 | 681 | 14 | CB690041 | CB690041 CBST-54-B |
| 22 | 177 | 7.9 | 675 | 14 | CF727824 | CF727824 UI-M-HB0- |
| 23 | 176 | 7.8 | 530 | 29 | CNS010PO | AL153820 Anopheles |
| 24 | 176 | 7.8 | 576 | 14 | CD295943 | CD295943 StrPu691. |
| 25 | 176 | 7.7 | 580 | 28 | BZ424995 | BZ424995 100023066 |
| 26 | 172.5 | 7.7 | 583 | 9 | AJ273947 | AJ273947 |
| 27 | 172.5 | 7.7 | 593 | 9 | AJ273918 | AJ273918 |
| 28 | 172.5 | 7.7 | 601 | 9 | AJ273921 | AJ273921 |
| 29 | 171.5 | 7.6 | 573 | 14 | CA937626 | CA937626 sav42b10. |
| 30 | 170.5 | 7.6 | 601 | 9 | AJ273050 | AJ273050 |
| 31 | 170.5 | 7.6 | 712 | 14 | CD311344 | CD311344 StrPu691. |
| 32 | 170.5 | 7.6 | 1029 | 29 | CNS071DW | AL424794 T7 end of |
| 33 | 170.5 | 7.6 | 895 | 13 | BQ216158 | BQ216158 AGENCOURT |
| 34 | 170 | 7.6 | 76 | 13 | BU575479 | BU575479 TGESTzyB8 |
| 35 | 169.5 | 7.5 | 545 | 13 | BU575479 | AJ273185 |
| 36 | 169.5 | 7.5 | 604 | 9 | AJ273185 | CF737198 UI-M-HD0- |
| 37 | 169 | 7.5 | 794 | 14 | CF737198 | AV107161 Zea mays |
| 38 | 169 | 7.5 | 1572 | 11 | AY107161 | BJ365857 |
| 39 | 168.5 | 7.5 | 601 | 12 | BJ365857 | AJ272712 |
| 40 | 168.5 | 7.5 | 609 | 9 | AJ272712 | BG246418 602360428 |
| 41 | 168.5 | 7.5 | 937 | 12 | BG246418 | CA510555 UI-R-FU0- |
| 42 | 168 | 7.5 | 831 | 14 | CA510555 | AL433387 T3 end of |
| 43 | 168 | 7.5 | 1050 | 29 | CNS0780L | AQ651427 Sheared D |
| 44 | 167.5 | 7.5 | 564 | 28 | AQ651427 | AJ274218 |
| 45 | 167.5 | 7.5 | 610 | 9 | AJ274218 | |

ALIGNMENTS

RESULT 1
BQ622771
LOCUS
DEFINITION
CC-Contig67 Conidiobolus cornatus ARSEF 512 Conidiobolus coronatus
CDNA, mRNA sequence.
BQ622771
VERSION
BQ622771.1 GI:21649940
EST.
KEYWORDS
Conidiobolus coronatus (Delacroixia coronata)
Conidiobolus coronatus
Eukaryota; Fungi; Zygomycota; Zygomycetes; Entomophthorales;
Ancylistaceae; Conidiobolus.
REFERENCE
1 (bases 1 to 1605)

| | |
|---------|--|
| AUTHORS | Freimoser, F. M., Screen, S., Hu, G. and St. Leger, R. J. |
| TITLE | ESR analysis of genes expressed by the zygomycete pathogen |
| JOURNAL | <i>Candidiobius coronatus</i> during optimized secretion of proteins |
| COMMENT | Unpublished (2002) |
| | Contact: Freimoser F. M. |
| | Department of Entomology |
| | University of Maryland |
| | 4112 Plant Sciences Building, College Park, MD 20742, USA |
| | Tel: 301 405 16 13 |
| | Fax: 301 314 92 90 |
| | Email: ff334@umail.umd.edu. |

| FEATURES | source |
|---|--------|
| Location/Qualifiers | |
| 1...1695 | |
| /organism="Conidiobolus coronatus" | |
| /mol_type="mRNA" | |
| /strain="ARSEF 512" | |
| /db_xref="taxon:34488" | |
| /clone_lib="Conidiobolus cornutus ARSEF 512" | |
| /note="vector: UniZap; Conidiobolus coronatus was grown in minimal medium supplemented with Manduca sexta cuticle and peptone for 18 hours. A cDNA library was constructed in the unidirectional Lambda vector UniZap." | |

ORIGIN

Alignment Scores:

| | | | |
|------------------------|----------|---------------|------|
| Pred. No.: | 1.57e-13 | Length: | 1605 |
| Score: | 246.00 | Matches: | 85 |
| Percent Similarity: | 43.99% | Conservative: | 43 |
| Best Local Similarity: | 29.21% | Mismatches: | 83 |
| Query Match: | 10.95% | Indels: | 80 |
| DB: | 13 | Gaps: | 14 |

US-09-985-689A-1-COPY (1-434) x B0622771 (1-1605)

```

US-09-985-699A-1-COPY (1-434) x BQ622771 (1-1605)

Qy 6 GlyIleValLysAlaAspValAlaGlnSerSer-----TyrGlyLeu----- 19
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
718 GGTGTGCTAAGGCTATTGCCGTCCAAGCCAATGCTCTTGGGGCTCTGTCGTGGTTGGT 777
Qy 20 -----TyrGlyGlnGly 23
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
778 CAACGTGCTAAGCTCGGATCTGCTCTTACTCTTACAACCATGCTGATGCTGGTCAAGGT 837
Qy 24 GlnIleValAlaValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMechHis 43
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
838 GTCACGTGTTTCGTTTTAGATACTGGTGTCAATGTCAGCCACAAATGAC----- 885
Qy 44 GluAlaPheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsnAspAla--- 62
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
886 -----TTCCGTGGTTCGTGCCACT-----TGGGGTACTTAACACTGCTGGT 924
Qy 63 -----AsnAspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsn 79
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
925 GGTAGCAACACTGATGCTCAGGTCACCGTACCGCTACTCACTGCTGTGTTACTATTTCGGT--- 98:
Qy 80 GlySerThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMet--- 98
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
982 -----ACCACCTATGGTGTGCCAAGAGGCTAACATGTTGCCGTTAAGGCTCTAGGT 1035
Qy 99 ---AspSerGlyGlyLeuGlyLeuProSerAsnLeuGlnThrLeuPheSerGln 117
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1036 GATGATGGCTCCGGATCATACTCTGGAATTATCTCCGGTATTGACTGGGTGTTAAG--- 1092
Qy 118 AlaTyrSerAlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAsnGlyAla 137
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1093 ---CACTCTGCTGCCAAGAAAGATTATCTATGACTTTAGAGGTTGGTAAAGCATGCT 1149
Qy 138 TyrThrTrpAspSerArgAsnValAspTyrValArgLysAsnAspMetThrIleLeu 157
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1150 CTTAACACT-----GCTGTTACAACGGCTGTCCAGCAGGGAGTTGCTCACTGTTGTC 1200
Qy 158 PheAlaAlaGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLys 177
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1201 ---GCTGCTGGTAAACGATAACAGAGATGCTGTGGATCTCT---CCCGCTCTGCTCCT 1254

```

| | | | |
|------------|--|---|-----------------|
| Qy | 178 | AsnAlaIleThrValGlyAlaThrGlu-----AsnLeuArgProSerPheGlySerTyr | 195 |
| Dd | 1255 | TCCGCCATTACCGTTGGTGCCCATGTCAATGATAAAAGGCCTTCATCTCTAAGTTC | 1314 |
| Qy | 196 | AlaAspAsnIleAsnHisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArg | 215 |
| Dd | 1315 | GGTAGCTGTGTC----- | 1326 |
| Qy | 216 | IleIysProaspValMetAlaProGLYThrPhelIleLeuSerAlaArgSerSerLeuAla | 235 |
| Dd | 1327 | -----GATATCTTAGTCTCTGGTGTCACAATCTCTCCACC----- | 1362 |
| Qy | 236 | ProAspSerSerPheTrpAlaAsnHisASPserLysTyralaTyMetGlyGlyThrSer | 255 |
| Dd | 1363 | -----TGGAAGGGATCTAACACTGCCACCAACCACTCTCTGGTACCTCT | 1407 |
| Qy | 256 | MetAlaThrPrIleValAlaGlyAsnValAla | 266 |
| Dd | 1408 | ATGCGTTGGCCTCACATIGCTGGTTAGTGTCT | 1440 |
| RESULT 2 | | | |
| BJ395336 | | | |
| LOCUS | BJ395336 | Dictyostelium discoideum cDNA library, SF Dictyostelium | EST 08-MAR-2002 |
| DEFINITION | BJ395336 | dictoideum cDNA clone dds38b16 5', mRNA sequence. | |
| ACCESSION | BJ395336 | | |
| VERSION | BJ395336.1 | GI:19306422 | |
| KEYWORDS | EST. | | |
| SOURCE | Dictyostelium discoideum | | |
| ORGANISM | Dictyostelium discoideum | | |
| REFERENCE | Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium. | | |
| AUTHORS | Urushinara,H., Tanaka,Y., Kohara,Y. and Shin-i,T. | | |
| TITLE | Full length cDNA of Dictyostelium discoideum at the slug stage | | |
| JOURNAL | Unpublished (2002) | | |
| COMMENT | Contact: Tadasu Shin-i | | |

Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tahini@genes.nig.ac.jp.

```

FEATURES
  source
    1. .540
      /organism="Dictyostelium discoideum"
      /mol_type="mRNA"
      /strain="AX4"
      /db_xref="taxon:44689"
      /clone="dds38b16"
      /sex="mat A"
      /dev_stages="slug stage"
      /clone_lib="Dictyostelium discoideum cDNA library, SF"

```

| ORIGIN | Alignment Scores: | Length: | 640 |
|------------------------|-------------------|---------------|-----|
| Pred. NO.: | 1.06e-13 | Matches: | 69 |
| Score: | 241.00 | Conservative: | 36 |
| Percent Similarity: | 49.16% | Mismatches: | 72 |
| Best Local Similarity: | 32.70% | Indels: | 34 |
| Query Match: | 10.73% | Gaps: | 9 |
| DB: | 12 | | |

US-09-985-689A-1-COPY (1-434) x BJ395336 (1-640)

QY 19 LeuTyGlyGlnGlyGlnIleValalavalalalaspThrGlyLeuaspThrGlyArg--- 37
 ::: ::: ::: ::: ::: :::
Db 22 TTAAGAGGTAAAGGTCAGATATTGCTGATCTGTTAGATGGTAGCCATTGT 81
 ::: ::: ::: ::: ::: :::
QY 38 -----AsnAspSer-----SerMetHisGluAlaPheArgGly 48
 ::: ::: ::: ::: ::: :::
Db 82 TTCCTTTTCAGATTCCAAGTATCCATATCAATTTAATCAAGTGAATGAAATCATGAAA 141

Qy 49 LysleThrAlaLeuTyrAlaLeuGlyArgThrAsnAsnAlaAsnAspThrAsnGlyHis 68
 Db 142 GTTGTACT-----TATATTACTACATGACATGAAGATTATGTAATGTCAT 192
 Qy 69 GlyThrHisValaGlySerValLeuGlyAsnGly-----Ser 81
 Db 193 GGTACACATGTTTGTGCTCTGCAGCAGGTACTCCAGAGGATTCTTTCATGGGCTATTTC 252
 Qy 82 ThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSerGly 101
 Db 253 TCATTAGTGGTCTTGCACTGACGCAAGATTGCAATTATGAT---CCTTCATCTGA 309
 Qy 102 GlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaIleSerAla 121
 Db 310 AGTTCTGAACCAACACACCCAGAGATACAGTCAAAATGATCAAAACCATATATGATGCA 369
 Qy 122 GlyAlaArgIleHisThrAsnSerTrpGlyAla-----AlaValAsnGlyAla 137
 Db 370 GGTGCAAGAGTACATGGTGATCTCTGGGGTCTGTATCTTTGCAAGGTTATATGGTGT 429
 Qy 138 TyrThrThrAspSerArgAsnValAspTyrValArgLys---AsnAspMetThrIle 156
 Db 430 TATTCGATGATGCTGGTGGTATTGATGATCTCTCTATGATGATGATGATGATGATGAT 489
 Qy 157 LeuPheAlaAlaGlyAsnGlyProAsnGlyGlyThrIleSerAlaProGlyThrAla 176
 Db 490 CTAAAGCTCTGCTGTAAT---AAGAGCTATTGTCATCTTTATGATCAAGCAACAGCT 546
 Qy 177 LysAsnAlaIleThrValGlyAlaThrGluAsnLeuArgProSerPheGlySer----- 194
 Db 547 AAAATGCAATTACAGTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 606
 Qy 195 -----TyrAlaAspAsnIleAsn 200
 Db 607 TTGGAATATTATGATTTCTCAGATAATGCTAAT 639
 RESULT 3
 AK029048
 LOCUS
 DEFINITION Mus musculus 10 days neonate skin cDNA, RIKEN full-length enriched library, clone:473248M11 product:membrane-bound transcription factor protease, site 1, full insert sequence.
 ACCESSION AK029048
 VERSION AK029048.1 GI:26325017
 KEYWORDS HTC; CAP trapper.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1
 Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
 TITLE High-efficiency full-length cDNA cloning
 JOURNAL Meth. Enzymol. 303, 19-44 (1999)
 MEDLINE 99279253
 PUBMED 10349636
 2
 Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
 TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
 JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
 MEDLINE 20499374
 PUBMED 11042159
 3
 Shibata,K., Itoh,M., Alizawa,K., Nagaoaka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kitsuai,T., Tashiro,H., Itoh,M., Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwara,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watanuki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
 RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer

JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 AUTHORS

Genome Res. 10 (11), 1757-1771 (2000)
 20530913
 11076861
 4
 The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium.
 Functional annotation of a full-length mouse cDNA collection
 Nature 409, 685-690 (2001)

JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 AUTHORS

The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team.
 Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs
 Nature 420, 563-573 (2002)
 6 (bases 1 to 4198)

JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 AUTHORS

Adachi,J., Alizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,M., Hayashida,K., Hayateu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T., Kato,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M., Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M., Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N., Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N., Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S., Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A., Muramatsu,M. and Hayashizaki,Y.
 Direct Submission

JOURNAL
 MEDLINE
 PUBMED
 REFERENCE
 AUTHORS

Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan [E-mail:genome-res@gsc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216]

COMMENT

cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
 Please visit our web site for further details.
 URL:http://genome.gsc.riken.go.jp/
 URL:http://fantom.gsc.riken.go.jp/
 Location/Qualifiers

FEATURES

source

1..4198
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="FANTOM:DB:473248M11"
 /db_xref="MGI:2391080"
 /db_xref="taxon:10090"
 /clone="473248M11"
 /tissue_type="skin"
 /clone_lib="RIKEN full-length enriched mouse cDNA library"
 /dev_stage="10 days neonate"
 417..3575
 /note="unnamed protein product; membrane-bound transcription factor protease, site 1 (MGD|MGI:1927235, GB|NM_019709, evidence: BLASTN, 99%, match=3782) putative"
 /codon_start=1
 /protein_id="BAC26263.1"
 /db_xref="GI:26325018"

CDS

VEFSTVVEYIYVAFNGYFTAKARNSEVENWRIIPRNPFSDYPSDFE
 VIQKEKQAGLTLEDHFNKVTQPKRFRSLKFAEENPVPFNTRQKQWSSR
 PLKRASLIGSGFWHATGRHSRRLLRPQVATQADVLQWGYTGANRVAVFD
 TGLSKPHFKNVKERTWNTERTLDDGLGHGTAVAGIASMRCQGFAPAEHLIFR
 VFTNNQSVTWFLEDAFNVALKKMDVLNLSIGGDFDNDHPDVKWELTANNVIMVS
 AICNDGPLYGLNPNADMDVIGGIDFENIAPSSRGMTWELPGYGRVKEDIV
 TYGAVRSGVGGRCGLRSVASPVVAGAVTLLVSTVQKRELNPASVQALIASA
 RLPGVNMFEQHGKLDLRLAYQLSSYQPAQSLSPYIDTECPNMPYCSQPIYIG
 GMPTIVNTILNGMGTGRIVDKPEWRPYLPQNGDNIEVAFSYSVLWFSWGLAIS

SVTKKAASWEGIAQGHIMITVASPAETELHSGAHTSTVKLPKIKVITPTPRSKRVL
 WQYNLRYPGVPFRDLNRMKNDPLDWNHGHVHTNFRDMYQHLRSMGYFVEVLGAPF
 TCEDATQGTGLLVDSEREPPEBIAKLRRDNDGLSLVIFSDWNTTSMKVKFYDE
 NTRQWMDPTGGANIPALNELLSVNMNPSDLGEBEFLANHDWYASGCSIAKFFE
 DQVITQPKDQGLSVLKQETAIVENVILGLYOIPSEGGSEIVLYGDSNCLDSDHQ
 KQCFMLLDALLOTYSYGVTPPSLSHSGNRQPPSAGLAPPERMEGHLHRYSKVLEA
 HGDGPKRPLPACPLHNPQPLNETAPSNLWKHKULLSIDLKVILFNPENRQPV
 RPLSPGSGANDIPGIMPRYNOBGVQIPVFAELGANVALAFVWQISKAKSRPKR
 RRPRAKRPQLAQCAHPARTPSV"
 4180..4185
 /note="putative"
 4198
 /note="putative"

polyA_signal

polyA_site

ORIGIN

Alignment Scores:

Pred. No.: 7,31e-11 Length: 4198
 Score: 226.50 Matches: 116
 Percent Similarity: 41.32% Conservative: 65
 Best Local Similarity: 26.48% Mismatches: 160
 Query Match: 10.08% Indels: 97
 DB: 11 Gaps: 25

US-09-985-689A-1-COPY (1-434) x AK029048 (1-4198)

Qy 8 ValylserAlaAspValAlaGlnSerSerTyGlyLeuTyGlyGlnGlyGlnIleValAla 27
 Db 1005 CTCAGGACGATGCTGGG---CAGATGGGATACACAGGTGCTTAATGTCAGATTGCT 1061
 Qy 28 ValAlaAspThrGlyLeuAspThrGlyArgAsnAspSerSerMetHisGluAlaPheArg 47
 Db 1062 GTTTTGTACTGGCTC-----AGTGAGAGCATCCGCAATTTTAAAG 1103
 Qy 48 GlyIysIleThrAlaLeuTyAlaLeuGlyArgThrAsnAsnAlaAsn----- 63
 Db 1104 AAT-----GTGAAGGAGAGAACCACTGGACCAATGAGCGGACCCCTG 1145
 Qy 64 ---AspThrAsnGlyHisGlyThrHisValAlaGlySerValIleGlyAsnGlySerThr 82
 Db 1146 GATGATGGGTAGGCCATGACATTCGTTGCAGGT---GTGATTCGACGATGAGGAG 1202
 Qy 83 AsnIysGlyMetAlaProGlnAlaAsnLeu---ValPheGlnSerIleMetAspSerGly 101
 Db 1203 TGCCAGGATTTGCTCCAGATGACAGCTGCACATCTTCAGGGTCTTTTACCACCAAT--- 1259
 Qy 102 GlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTySerAla 121
 Db 1260 -----CAGGTGCTTACACATCTTGGTTCGATGCGCTCAACTAT 1301
 Qy 122 GlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAsnGlyAlaTyThrThrAsp 141
 Db 1302 GCCATCCTAAGAAGATGGAGCTTCTCACTTAGCATCGTGGGCCGCACTTCATGGAT 1361
 Qy 142 SerArgAsnValAlaAspTyTyr-----ValArgIysAsnAspMetThrIleLeuPheAla 159
 Db 1362 CATCCGTTTGTGCAAGGTGGGAATTAACAGCTAACATGTAATTAATGTTTCTGCT 1421
 Qy 160 AlaGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaIysAsnAla 179
 Db 1422 ATTGGCAATGATGGACCTCTCTATGGCACTCTGAATAACCCCTGTCATCAGATGGATGTG 1481
 Qy 180 IleThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyAlaAspAsnIle 199
 Db 1482 ATTGGATGGGTGGCATTGAC-----TTTGAGATACATC 1517
 Qy 200 AsnHisValAlaGlnPheSerSerArgGly-----ProThrIysAsp 213
 Db 1518 -----GCTCGCTTTCTTCAGGGGAATGACTACTCTGGGAATTAACAGGAGGCTAT 1568
 Qy 214 GlyArgIleIysProAspValMetAlaProGlyThrPheIleLeuSerAlaArgSerSer 233
 Db 1569 GGTGCTGTGAAGCTGACATTGTC-----ACCTATGGTGTGAGTGGCGGGTTC 1619

Qy 234 LeuAlaProAspSerSerPheTrpAlaAsnHisAspSerIysTyAlaTyMetGlyGly 253
 Db 1620 GGTGTGAAAGGGGGCTGC-----CGTGCACCTCTCAGG 1652
 Qy 254 ThrSerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPheVal 273
 Db 1653 ACAGTGTGCTTCCCACTGGTGTGCTGGGCGGTCACTTGTAGTAGCAGCAGTACAG 1712
 Qy 274 LysAsnArgGlyIleThrProIysProSerLeuLeuIysAlaAlaLeuIleAlaGlyAla 293
 Db 1713 AAGCGGAGCTGGTG-----AATCTGCCAGTGTGAAGCAAGCTTTGATAGCGTCAGCC 1766
 Qy 294 AlaAspIleGlyLeuGlyTyProAsnGlyAsn-----GlnGlyTrpGlyArgVal 310
 Db 1767 CGGAGACTT-----CCTGGGTCAACATGCTCGAGCAAGTCAATGCAAGTTG 1814
 Qy 311 ThrIleAspIysSerLeuAsnVal---AlaTyValAsnGluSerSerSerLeuSerThr 329
 Db 1815 GATCTGCTGGAGCTTATCAGATCTCAGCAGCTATAAACCGCAGCAGCCTGAGTCCT 1874
 Qy 330 Ser-----GlnIysAlaThrTySerPheThrAlaThrAlaGlyLysProLeu 345
 Db 1875 AGCTACATCAGCAGTACTGCTGCTCCTACATGTTG---CCCTACTGCTCCAGCCT--- 1928
 Qy 346 LysIleSerLeuValTrpSerAspAlaProAlaSerThrThrAlaSerValThrLeuVal 365
 Db 1929 -----ATCTACTATGAGGAATGCCA-----ACAATCGTTAATGTCAACATCTCT 1973
 Qy 366 AsnAspIleAsnLeu-----ValIleThrAlaProAsnGlyThrGlnTyValGly 382
 Db 1974 AATGCGATGGCGTCAAGGAAGAATTTGGTAAGCTGAGTGGCGGACCTATTATTA--- 2030
 Qy 383 AsnAspPheThrSerProTyAsnAspAsnTrpAspGlyArgAsnAsnValGluAsnVal 402
 Db 2031 -----CCACAGAAATGGA-----GACACATTAAGTGAAGTGGCC 2060
 Qy 403 Phe-----IleAsnAlaProGlnSerGlyThrTyThrIleGluVal 416
 Db 2061 TTCTCTACTCCTCAGTGTGTTGGCCCTGCTGCTGCTTACCTTGCATCTCCATT 2114
 RESULT 4
 BU393752 594 bp mRNA linear EST 08-MAR-2002
 LOCUS BU393752 Dictyostelium discoideum cDNA library, SF Dictyostelium
 DEFINITION discoideum cDNA clone dds32b16 5', mRNA sequence.
 BU393752
 ACCESSION BU393752.1 GI:19304838
 VERSION BU393752
 KEYWORDS EST.
 SOURCE Dictyostelium discoideum
 ORGANISM Dictyostelium discoideum
 Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
 REFERENCE 1 (bases 1 to 594)
 AUTHORS Urushihara,H., Tanaka,Y., Kohara,Y. and Shin-i,T.
 TITLE Full length cDNA of Dictyostelium discoideum at the slug stage
 JOURNAL Unpublished (2002)
 COMMENT Contact: Tadasu Shin-i
 Center For Genetic Resource Information
 National Institute of Genetics
 1111 Yata, Mishima, Shizuoka 411-8540, Japan
 Tel: 81-559-81-6856
 Fax: 81-559-81-6855
 Email: tshini@genes.nig.ac.jp.
 Location/Qualifiers
 1..594
 /organism="Dictyostelium discoideum"
 /mol_type="mRNA"
 /strain="AX4"
 /db_xref="taxon:44689"
 /clone="dds32b16"
 /sex="mat A"
 /dev_stage="Slug stage"
 /clone_lib="Dictyostelium discoideum cDNA library, SF"

FEATURES source

ORIGIN

Rockville, MD. Genomic DNA isolated from a cloned population of trypanosome. Brucsize distribution (CPGUS27/GMAT 10.1) was mechanically sheared to give a tight, narrow distribution (CPGUS27/GMAT 10.1). The library construction is described in detail in Smith, H. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Barrell. Oxford University Press, 1999).

Email: nelsayed@tigr.org
Details of T. brucei sequencing at the Sanger Centre are available at <http://www.sanger.ac.uk/projects/Tbrucei/>.

FEATURES

Source

```
1. 532
/organism="Trypanosoma brucei"
/mol_type="genomic DNA"
/strain="TREU927"
/db_xref="taxon:5691"
/clone="319q10"
```

ORIGIN

| | | |
|------------------------|----------|------------------|
| Alignment Scores: | | |
| Pred. No.: | 1.44e-11 | 532 |
| Score: | 218.50 | 61 |
| Percent Similarity: | 48.3% | Conservative: 68 |
| Best Local Similarity: | 33.1% | Mismatches: 28 |
| Query Match: | 9.72% | Indels: 27 |
| DB: | 29 | Gaps: 6 |

U9-09-985-689A-1-COPY (1-434) X TA319G10P (1-532)

| | | | |
|-----|----|---|-----|
| 93 | Qy | ValPheGlnSerIleMetAspSerGlyGlyGlyLeuGly----- | 105 |
| 2 | Db | GTAATGCCAAGTAAATGGTTCGCCACGGGGGCAAGATTCTCAGGGTGGGCTGCC | 61 |
| 106 | Qy | -----GlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyr | 119 |
| 62 | Db | CATCCAGCTCAGAGCTGTCTCCCCACAGACGTTACTCAAATTATTCGTCGCGGTATAT | 121 |
| 120 | Qy | SerAlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAsnGlyAlaTyrThr | 139 |
| 122 | Db | GGCGCTGGAGCCCGTGTCTCTAAACTCGTGGGGTTTTGTCTCCCTCCCGAGATTCT | 181 |
| 140 | Qy | ThrAspSerArgAsnValAspAspTyrValArgLys---AsnAspMetThrIleLeuPhe | 158 |
| 182 | Db | GCTGTGGAAAGGATATGGATGAGTTTGCAGTAGTATGACATCGCTACTTATCTTC | 241 |
| 159 | Qy | AlaAlaGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsn | 178 |
| 242 | Db | TCCACTGGCAACAGTATCCAAAGATGCG-----CTAATGACTCGGTGCGGTGAAGAC | 295 |
| 179 | Qy | AlaIleThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsn | 198 |
| 296 | Db | GTGATGTGCGTGGGGTCCACAAAAGCGT-----TTTGACGCTTCGAAGAC--- | 343 |
| 199 | Qy | IleAsnHisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysPro | 218 |
| 344 | Db | -----ATTGTTCTTCGTTTCTTCGCATGGTCCAAACATCACGACGCTAGGATGAACCC | 397 |
| 219 | Qy | AspValMetAlaProGlyThrPheIleLeuSerAlaArgSerSerIleuAlaProaspSer | 238 |
| 398 | Db | GATCTGTGCGTCCCGGGAAGAGGTGTGCTCTTTCTTCCTGGCAAGCATCAGT | 457 |
| 239 | Qy | SerPheTrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyThrSerMetAlaThr | 258 |
| 459 | Db | -----CAATGTTAAAGTGTGGCCCAAGCGGGNGTTTCATCGTGGCAACT | 502 |

259 ProfilValAla 262

503 GGGGCTGGG 514

RESULT 6

BQ142519/c

LOCUS

2141 bp mRNA

linear

EST 24-APR-2002

| Alignment Scores: | |
|------------------------|---------|
| Pred. No.: | 4,3e-12 |
| Score: | 224.50 |
| Percent Similarity: | 50.80% |
| Best Local Similarity: | 31.55% |
| Query Match: | 9.99% |
| DB: | 12 |
| Length: | 594 |
| Matches: | 59 |
| Conservative: | 36 |
| Mismatches: | 59 |
| Indels: | 33 |
| Gaps: | 9 |

US-09-985-689A-1-COPY (1-434) x BJ393752 (1-594)

| | | | |
|----|-----|--|-----|
| Qy | 49 | IyslleThrAlaLeuTyAlaLeuGlyArgThrAsnAsnAlaAsnAspThrAsnGlyHis | 68 |
| | | | |
| Db | 21 | AAAGTTGAACCTATATTACCACATCAACACGACGATAGTAAAGTCAGTGGTCAC | 80 |
| | | | |
| Qy | 69 | GlyThrHisValAlaGlySerValLeuGly | 81 |
| | | | |
| Db | 81 | GGTACACATATTGTGGTCTGCGACGAGTACTCCAGAGGATTCCTCAGTTAAATTTC | 140 |
| | | | |
| Qy | 82 | ThrAsnLysGlyMetAlaProGlnAlaAsnLeuValPheGlnSerIleMetAspSerGly | 101 |
| | | | |
| Db | 141 | TCATTATAGTGGTCTTGCAACTGATGCAAGATTGTCATTC | 191 |
| | | | |
| Qy | 102 | GlyGlyLeuGlyGlyLeu | 119 |
| | | | |
| Db | 192 | AGTGGTTTCATCAAGTTTCACACCTCCATCGGATTTGAACAATTATATCAACCATTTAT | 251 |
| | | | |
| Qy | 120 | SerAlaGlyAlaArgIleHisThrAsnSerTrpGlyVala | 135 |
| | | | |
| Db | 252 | GACGCAGGTGCAAGAGTGATGTGATTCITGGGGTTCGTATCAGTAGGGGTATACA | 311 |
| | | | |
| Qy | 136 | GlyAlaTyThrThrAspSerArgAsnValAspAspTyValArgLysAsn | 154 |
| | | | |
| Db | 312 | GGTAGTTATCATCAGACACTGCTTCATTCATGATTCCTTTCTCACTCATCCAGATTTC | 371 |
| | | | |
| Qy | 155 | ThrIleLeuPheAlaAlaGlyAsnGluGlyProAsnGlyGlyThrIleSer | 172 |
| | | | |
| Db | 372 | ATCATTTCTTAGAGTCTGGTAAC | 422 |
| | | | |
| Qy | 173 | ProGlyThrAlaLysAsnAlaIleThrValGlyAlaThrGluAsnLeuArg | 189 |
| | | | |
| Db | 423 | CAATCCACTGCAAGAAAGTTTATTACCGTTGGTGCTCATCAACAATTCATGAAATTTAT | 482 |
| | | | |
| Qy | 190 | -----ProSerPheGlySerTyrrAlaAspAsnIleAsn | 201 |
| | | | |
| Db | 483 | TTAACTGATGGTCCAAATTATATAATTATCAATCACTGTCGATATAAATCAAGATTTA | 542 |
| | | | |
| Qy | 202 | ValAlaGlnPheSerSerArg | 208 |
| | | | |
| Db | 543 | ATATGTGATTTCATAGCAGA | 563 |
| | | | |

| LOCUS | TA319G10P | 532 bp | DNA | linear | GSS 13-DEC-2000 |
|------------|--|--------|-----|--------|-----------------|
| DEFINITION | T. brucei sheared genomic DNA clone 319g10, forward sequence, genomic survey sequence. | | | | |

ACCESSION
AL492464

VERSION AL492464.1 GI:11867408

KEYWORDS
SOURCE
GSS.
TRANSDOME
PRICE;

| SOURCE | ORGANISM |
|--------------------|--------------------|
| Trypanosoma brucei | Trypanosoma brucei |
| Trypanosoma brucei | Trypanosoma brucei |

Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;

Trypanosoma.

REFERENCE
I (bases 1 to 532)
WALL N. Bowman S. Lennard N. J. Daggett J. Atkin R.

Authors: Hail, N., Bowman, S., Lehman, R. C., DeGroot, J., Melville, S. E., Chillingworth, C., Ormond, D., Harris, B., El-Sayed, N., Hou, L., Melville, S. E., Ratandorn, M. A. and Barrell, B. G.

| TITLE | Direct Submission |
|-------|-------------------|
| | |

JOURNAL Submitted (10-DEC-2000) Trypanosoma brucei genome sequencing

project, Sanger Centre, The Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA. E-mail: barrell@sanger.ac.uk and

Cambridge CB10 1SA, E-mail: patrick.banger@ac.uk and
nhl@sanger.ac.uk

COMMENT

| | | | |
|-------------|-----|--|-----|
| QY | 156 | lleLeuPheAlaAlaGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGlyThr | 175 |
| Db | 681 | CTTGCCGTCGCGCTGGCAACGATTAACCGGATGCCCAAAACACCTCT--CCCGCTTCC | 625 |
| QY | 176 | AlaLysAsnAlaIleThrValGlyAlaThr-----GluAsnLeuArgProSerPheGly | 193 |
| Db | 624 | GAGCCCTACTGCTGCACCTGTTGGTCCCACTCGTCAGATGACAGCCGATCTACCTTTTCC | 565 |
| QY | 194 | SerTyrAlaAspAsnIleAsnHisValAlaGlnPheSerSerArgGlyProThrLysAsp | 213 |
| Db | 564 | AACTAC----- | 559 |
| QY | 214 | GlyArgIleLysProAspValMetAlaProGlyThrPheIleLeuSerAlaArgSerSer | 233 |
| Db | 558 | GCACAGATT--GTCGATATCTTCCTCTCTGTACCGCATCTCTCCACC-- | 511 |
| QY | 234 | LeuAlaProAspSerSerPheTyrAlaAsnHisAspSerLysTyrAlaTyrMetGlyGly | 253 |
| Db | 510 | -----TGGATCAAT-----GGCCGCCACCAACACCATCTCTGGC | 478 |
| QY | 254 | ThrSerMetAlaThrProIleValAlaGlyAsnValAla | 266 |
| Db | 477 | ACCTCCATGGCTACTCCCATATTGCTGCTTGTCTGTC | 439 |
| RESULT 7 | | | |
| ACQ552212/c | | | |
| LOCUS | | | |
| DEFINITION | | | |
| ACCESSION | | | |
| VERSION | | | |
| KEYWORDS | | | |
| SOURCE | | | |
| ORGANISM | | | |
| REFERENCE | | | |
| AUTHORS | | | |
| TITLE | | | |
| JOURNAL | | | |
| COMMENT | | | |
| FEATURES | | | |
| source | | | |

http://image.llnl.gov
This clone was contributed by the Brain Molecular Anatomy Project
(BMAP)

Seq primer: pYX-5.

FEATURES

source

Location/Qualifiers
1..771
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6"
/db_xref="taxon:10090"
/clone="IMAGE:6816072"
/tissue_type="whole brain"
/dev_stage="embryo 13.5 14.5 16.5 17.5 dpc"
/lab_host="DH10B (T1 phage resistant)"
/clone_lib="NIMH_BMAP_FMO"
/note="Organ: Brain; Vector: pYX-Asc; Site: 1: EcoR I; Site 2: Not I; The library was constructed according to Bonafo, Lennon and Soares, Genome Research, 6:791-806, 1996. Denatured RNA was size fractionated on a 1% agarose gel. First strand cDNA synthesis was primed with oligo-dT primer containing a Not I site. Double strand cDNA was size selected according to mRNA size fraction, ligated with EcoR I adaptor, digested with NotI and then cloned directionally into pYX-Asc vector. The library tag sequence located between the Not I site and the polyA tail is AGCGAGACAG. This library was created for the University of Iowa Brain Anatomy Project (BMAP). 'Gene Discovery in the Developing Mouse Nervous System', supported by National Institute of Mental Health (NIMH), Hemlin Chin, Ph.D., program coordinator."

ORIGIN

Alignment Scores:
Pred. No.: 6,26e-09 Length: 771
Score: 195.00 Matches: 82
Percent Similarity: 41.81% Conservative: 38
Best Local Similarity: 28.57% Mismatches: 109
Query Match: 8.68% Indels: 58
DB: 14 Gaps: 13

US-09-985-689a-1-COPY (1-434) x CA320325 (1-771)

Qy 8 vallyalaaspvalalaglnserstyrglyleutyrglyglnlevalala 27
Db 40 CTCAGGCAGATGCTGTGG-----CAGATGGGATACACAGGTGCTAATCTCAGAGTGTCT 96
Qy 28 valalaaspthrglyleuaspthrglyargasnaspsersermethleuhalphearg 47
Db 97 GTTTTGTACTGGCTC-----AGTGAGAGCATCCGCATTTTAAG 138
Qy 48 GlylylethrAlaLeuTyRAlaLeuGlyArgThrAsnAlaAsn----- 63
Db 139 AAT-----GTGAAGGAGAGAACCACTGGACCAATGAGCGGACCCCTG 180
Qy 64 ---AspThrAsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGlySerThr 82
Db 181 GATGATGGGTAGCCCATGGCATTCGTTGCAAGT---GTGATCCAGCATGAGGGAG 237
Qy 83 AsnlysglyMetAlaProGlnAlaAsnLeu---ValPheGlnSerIleMetAspSerGly 101
Db 238 TGCCAGAGATTGCTCCAGATGAGAGCTGCACATCTTCAGGGTCTTTACCAACAAT--- 294
Qy 102 GlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyRSerAla 121
Db 295 -----CAGGTGCTTACACATCTTGGTTTCTGGATGCTTCCTCAACTAT 336
Qy 122 GlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAsnGlyAlaTyRThrAsp 141
Db 337 GCCATCTAAAGAGATGGAGCTTCTCACTAGCATCGTGGCGCGCATTCATGGAT 396
Qy 142 SerArgAsnValAspAspTyR-----ValArglyAsnAspMetThrIleLeuPheAla 159
Db 397 CATCCGTTTGTGATGAGGTGGGAATTAAACAGCTAACCAATGTAATTATGGTTTCTGCT 456

Qy 160 AlaGlyAsnGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsnAla 179
Db 457 ATTGGCAATGATGGACCTCTATGGCACTCTGAATAACCTGCTGATCAGATGGATGG 516
Qy 180 IleThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyRAlaAspAsnIle 199
Db 517 ATTGGAGTGGTGGCATTGAC-----TTTGAAGATAACATC 552
Qy 200 AsnHisValAlaGlnPheSerSerArgGly-----ProThrLysAsp 213
Db 553 -----GCTCGCTTTCTTCAGGGGAATGACTACCTGGGAATTACAGAGGCTAT 603
Qy 214 GlyArgIleLysProAspValMetAlaProGlyThrPheIleLeuSerAlaArgSerSer 233
Db 604 GTCGTGTGAGGCTGACATTGTC-----ACCTATGGTGTCTGAGTGGCGGTTC 654
Qy 234 LeuAlaProAspSerSerPheTrpAlaAsnHisAspSerLysTyRAlaTyRMetGlyGly 253
Db 655 GGTGTGAAGGGGGCTGC-----CGTGCACCTCTCAGGG 697
Qy 254 ThrSerMetAlaThrProIleValAlaGlyAsnValAlaGlnLeuArgGluHisPheVal 273
Db 688 ACCAGTGTCTTCCCGAGGTGCTGGGGGCTCACCTTGTATTGTANGCACAGATACAG 747
Qy 274 LysAsnArgGlyIleThrPro 280
Db 748 AAGCGGAGGCTGTGAATCCT 768

RESULT 11

BU369190 633 bp mRNA linear EST 08-MAR-2002
LOCUS BU369190 Dictyostelium discoideum cDNA library, CF Dictyostelium
DEFINITION BU369190 Dictyostelium cDNA clone ddc49116 5', mRNA sequence.
ACCESSION BU369190
VERSION BU369190.1 GI:19278573
KEYWORDS EST.
SOURCE Dictyostelium discoideum
ORGANISM Dictyostelium discoideum
REFERENCE 1 (bases 1 to 633)
AUTHORS Urushihara, H., Tanaka, Y., Kohara, Y. and Shin-i, T.
TITLE Full length cDNA of Dictyostelium discoideum at the culmination stage
JOURNAL Unpublished (2002)
COMMENT Contact: Tadasu Shin-i
Center For Genetic Resource Information
National Institute of Genetics
1111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshini@gene.nig.ac.jp.
Location/Qualifiers
1..633
/organism="Dictyostelium discoideum"
/mol_type="mRNA"
/strain="AX4"
/db_xref="taxon:44689"
/clone="ddc49116"
/sex="mat A"
/dev_stage="Culmination stage"
/clone_lib="Dictyostelium discoideum cDNA library, CF"

FEATURES

source

1..633
/organism="Dictyostelium discoideum"
/mol_type="mRNA"
/strain="AX4"
/db_xref="taxon:44689"
/clone="ddc49116"
/sex="mat A"
/dev_stage="Culmination stage"
/clone_lib="Dictyostelium discoideum cDNA library, CF"

ORIGIN

Alignment Scores:
Pred. No.: 1,14e-08 Length: 633
Score: 191.00 Matches: 57
Percent Similarity: 41.55% Conservative: 29
Best Local Similarity: 27.54% Mismatches: 49
Query Match: 8.50% Indels: 72
DB: 12 Gaps: 6

US-09-985-689a-1-COPY (1-434) x BU369190 (1-633)

Qy 136 GlyAlaTyrThrAspSerArgAsnValAspAspTyrValArgLys--AsnAspMet 154
 Db 6 GGTGGTATTCCGATGCTGGTGGTATTGATGCAATCCCTTATGATACCCAGAAATC 65
 Qy 155 ThrLeuPheAlaAlaGlyAsnGluGlyProAsnGlyGlyThrIleSerAlaProGly 174
 Db 66 TCTATCTAAGAGCTGCTGGTAAT---AACGAGCTATTGGCATCTTTATTAGCTCAAGCA 122
 Qy 175 ThrAlaLysAsnAlaIleThrValGlyAlaThrGluAsnLeuArgProSerPheGlySer 194
 Db 123 ACAGCTAAATGCAATACATGCTGGTCTGGAGCAACAGCTCATGTAATATATGTGCA 182
 Qy 195 -----TyrAlaAspAsnIle----- 199
 Db 183 GATGCAATGGAATATTGATGTTCTCAGATATGCTAATTTTCAAAGACCATGTTATTC 242
 Qy 199 ----- 199
 Db 243 GATAAGAGCTATTGTAATATTATAGACCGCTAAATGCTGTCAGAGTTTCAAATGTTAA 302
 Qy 199 ----- 199
 Db 303 GGTTCACAATTATGTTGTCAGCATCTATTAAACAAATGCGATTCGATTATTCAACA 362
 Qy 200 -----AsnHisValAlaGlnPheSerSerArgGlyProThrLys 212
 Db 363 CAACCTCAATTTATATGAAATATATGGGATCATCTCAACAAAGGTCACACAT 422
 Qy 213 AspGlyArgIleLysProAspValMetAlaProGlyThrPheIleLeuSerAlaArgSer 232
 Db 423 GATGAGTAGATTGAACCTGATATAGTTCACCTGGTGAATATATACATCGGAGATCA 482
 Qy 233 Ser-----LeuAlaProAspSerSerPheTrpAlaAsnHisAspSer 246
 Db 483 AATGTTGAGAAATTCACAGCAATGTTGGTGTCTCTTTA-----CCAAATGCCAAT 536
 Qy 247 LysTyrAlaTyrMetGlyGlyThrSerMetAlaThrProIleValAlaGlyAsnValAla 266
 Db 537 GGTCTAATGTCATATCTGGTATCATCATGGCAACACCATTCGGACAGCAGCACACA 596
 Qy 267 GlnLeuArgGluHisPheVal 273
 Db 597 ATTCTTAGACAATATTAGTT 617

RESULT 12

BI750157 718 bp mRNA linear EST 25-SEP-2001
 LOCUS Fg02_10508_R Fg02_AAPC_EORC_Fusarium_graminearum_mycelium
 DEFINITION Gibberella zeae cDNA clone Fg02_10508, mRNA sequence.

ACCESSION BI750157

VERSION BI750157.1 GI:15771959

KEYWORDS EST

SOURCE Gibberella zeae

ORGANISM Gibberella zeae

REFERENCE Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;

Hypocreomycetidae; Hypocreales; Nectriaceae; Gibberella.

1 (bases 1 to 718)

Harris, L.J., Glassco, T., Rocheleau, H., Allard, S., Chapados, J.,

Couroux, P., De Moers, A., Hattori, J.I., Ouellet, T., Robert, L.S.,

Singh, J.A., Spott, D., and Tinker, N.A.

Expressed Sequence Tags from Fusarium graminearum mycelium

Unpublished (2001)

Contact: Harris, Linda J.

Eastern Cereal and Oilseed Research Centre

Agriculture and Agri-food Canada

Bldg. 21, Central Experimental Farm, Ottawa, Ontario, K1A 0C6,

CANADA

Tel: (613) 759-1314

Fax: (613) 759-6566

Email: harris.lj@em.agr.ca.

Location/Qualifiers

1. 718

FEATURES
 source

/organism="Gibberella zeae"
 /mol_type="mRNA"
 /strain="DAOM 180378"
 /db_xref="taxon:5518"
 /clone="Fg02_10508"
 /tissue_type="Mycelial tissue"
 /dev_stage="Asexual"
 /lab_host="E. coli (Sure cells)"
 /clone_lib="Fg02_AAPC_EORC_Fusarium_graminearum_mycelium"
 /notes="Vector: Bluescript SK+/XhoI-EcoRI; Site 1: EcoRI;
 Site 2: XhoI; Mycelial tissue was collected from V8 agar
 plates after a growth period of 6-7 days at 25 C with 14
 hrs (FL/UV) day ligh exposure. Mycelia was ground in
 liquid nitrogen prior it's storage at -80 C until RNA
 extraction. Directional cloning with 5' end of cDNA cloned
 into EcoRI site of pBluescript and 3' end of cDNA cloned
 into XhoI site of pBluescript (Stratagene, La Jolla, CA)."

ORIGIN

Alignment Scores:
 Pred. No.: 1.77e-08 Length: 718
 Score: 190.00 Matches: 66
 Percent Similarity: 44.05% Conservatives: 34
 Best Local Similarity: 29.07% Mismatches: 81
 Query Match: 8.46% Indels: 46
 DB: 12 Gaps: 10
 US-09-985-689A-1-COPY (1-434) x BI750157 (1-718)
 Qy 46 PheArgGlyLysIleThrAlaLeuTyrAlaLeuGlyArgThrAsnAsnAlaAspThr 65
 Db 9 TTCGAGGTCGTCTCTCAGGCTGTCTACACTGCTTCAGCGGGCCAGAACGCT---GACACC 65
 Qy 66 AsnGlyHisGlyThrHisValAlaGlySerValLeuGlyAsnGlySerThrAsnLysGly 85
 Db 66 AAGGTTCACGAAGCACTCAGTTGCTGCACACTATTGCCGGAAG-----ACATACGCT 116
 Qy 86 MetAlaProGlnAlaAsnLeu-----ValPheGlnSerIleMetAspSerGly 101
 Db 117 GTTGCCAAAGAAAGGCCACCATCCCAAGCTGTCAAGGTCTTCAG----- 158
 Qy 102 GlyGlyLeuGlyGlyLeuProSerAsnLeuGlnThrLeuPheSerGlnAlaTyrSer--- 120
 Db 159 ---GGTAGTTTCATCCAGCACTCCATCATCTCGCTGGCTTCACTGGGCTGCAACAGCAG 215
 Qy 121 ----AlaGlyAlaArgIleHisThrAsnSerTrpGlyAlaAlaValAsnGlyAlaTyr 138
 Db 216 ATCATCTCCAAAGGCCGACCAAGACACTCATATATGCTCTCGCGGTGGTTAC 275
 Qy 139 ThrThrAspSerArgAsnValAspAspTyrValArgLysAsnAspMetThrIleLeuPhe 158
 Db 276 TCTGCTTCTCTTCAACACACGCTGTCGAGCTCTKCTCCAGCTCCGGTATTATCTCTGCCA 335
 Qy 159 AlaAlaGlyAsnGlyProAsnGlyGlyThrIleSerAlaProGlyThrAlaLysAsn 178
 Db 336 GCTGCCGGTAACGATGTCGCAACGCTGCCAACACTTCT---CTGCTCTTCTCCAGC 392
 Qy 179 AlaIleThrValGlyAlaThrGluAsnLeuArgProSerPheGlySerTyrAlaAspAsn 198
 Db 393 GCCAWCACTGTCGGTGCATTGACAGC-----AAGTGGGC----- 428
 Qy 199 IleAsnHisValAlaGlnPheSerSerArgGlyProThrLysAspGlyArgIleLysPro 218
 Db 429 -----ATTCKCTCGTACTCAACTACGTTACCGTTCTC----- 461
 Qy 219 AspValMetAlaProGlyThrPheIleLeuSerAlaArgSerSerLeuAlaProAspSer 238
 Db 462 GATATCTTGGTCCTKGCACGAGGTTCTCTCGGC----- 497
 Qy 239 SerPheTrpAlaAsnHisAspSerLysTyrAlaTyrMetGlyGlyThrSerMetAlaThr 258
 Db 498 -----TGGTACACCAACAGCAGTGCACCAACACCATCAGCGKACSTCCNTGGTACT 551

| | | |
|---|----------|--|
| QY | 259 | ProfileValalaGlyAsnVal 265 |
| Db | 552 | CYCCACATTGGCGACTTGTC 572 |
| RESULT 13 | | |
| BC060627 | | |
| LOCUS | | |
| DEFINITION | | Mus musculus tripeptidyl peptidase II, mRNA (cDNA clone IMAGE:6844459), containing frame-shift errors. |
| ACCESSION | | BC060627 |
| VERSION | | |
| KEYWORDS | | HTC |
| SOURCE | | BC060627.1 GI:38148645 |
| ORGANISM | | Mus musculus (house mouse) |
| REFERENCE | | |
| AUTHORS | | |
| FEATURES | | |
| source | | |
| db_xref= | | "taxon":10090" |
| /clone= | | "IMAGE:6844459" |
| /tissue_type= | | "Brain, mouse, 13.5,14.5,16.5,17.5 dpc" |
| /clone_lib= | | "NIH_BMAP_FY0" |
| /lab_host= | | "DH10B" |
| /notes= | | "Vector: pYX-ASC" |
| ORIGIN | | |
| Alignment Scores: | | |
| Pred. NO.: | 9.36e-07 | Length: 4662 |
| Score: | 186.50 | Matches: 101 |
| Percent Similarity: | 35.77% | Conservative: 70 |
| Best Local Similarity: | 21.13% | Mismatches: 154 |
| Query Match: | 8.30% | Indels: 153 |
| DB: | 11 | Gaps: 21 |
| US-09-985-689A-1-COPY (1-434) x BC060627 (1-4662) | | |
| QY | 10 | AlaaspValalaglnSerSertyrGlyLeutyrglynglgnlevalavalala 29 |
| Db | 803 | GCTGAGATGCTGAATTACTCTGTGAACAATTTATGACGATGGGAACCTGCTTCCATTGTG 862 |
| QY | 30 | AspthrcglyLeuspthrGlyArgasnAspSerMetHisglualapheargglylys 49 |
| Db | 863 | ACCAGCCGGA----- 871 |
| QY | 50 | IlethrAlalaLeutyralaLeuGlyArgrThrAsnAsnAlaasnAspThrasnGlyHisgly 69 |
| Db | 872 | -----GGAGCTCATGCA 883 |
| QY | 70 | ThrhivallalaglyserValleuGlyAsn-----GlySerThrasnlysglyMet 86 |
| Db | 884 | ACCATGATGCAAGTAGTATAGCGCGAGGCATTTTCAGAGAAGCCTTGAAACGGAATGGAGTT 943 |
| QY | 87 | AlaproGlnAlaasnLeuValPhe-----GlnSerIleMetAsp 99 |
| Db | 944 | GCTCTGTGCTCAAATTCCTATCCATTAGATGGTGATACACGGCTAGACACTATGGAA 1003 |
| QY | 100 | SerGlyglyLeuGlyLeuProSerAsnLeuGlnThrleuPheSerGlnala--- 118 |
| Db | 1004 | ACAGCACAGGCCCTC-----ATCAGAGCTATGATAAGAGTTTATAATCATAGGTGTGAT 1057 |
| QY | 119 | -----TyrsAlaGlyAlaAtgIleHisThrAsnSerTrpGlyAlaalaValasn 135 |
| Db | 1058 | CTTGTCACATACAGTTATGGAGAGCACTCATTCGCCAAATTCCTGGAGAAATT----- 1111 |
| QY | 136 | GlyAlatyrThrThrAspSerArgasnValasPasytyrValarglyAsnAspMetThr 155 |
| Db | 1112 | -----TGTGAAGTAATTAATGAAGCAGTATGGAACATATAATCAATT 1153 |
| QY | 156 | IleleuPheAlaAlaGlyAsnGluGlyProasnGlyGlyThrileSerAlaPro---Gly 174 |
| Db | 1154 | TATGTTTCAAGTGTGGAATAATATGGTCCATGCTTTCTACAGTGGGTGTCCAGAGAGA 1213 |
| QY | 175 | ThrAlalyAsnAlaIleThrValGlyAlaThrGluAsnLeuArgProSerPheGlySer 194 |
| Db | 1214 | ACTACATCCAGTGTGATAGTGTGGAGCT----- 1243 |
| QY | 195 | TyrAlaAspAsnIleasnHisValAlaGlnPhe----- 205 |
| Db | 1244 | TATGTTTTCCCCTGATATATGATGTTGTCAGAGTATTCACGTAGAGAGAAACCTGCCTCGAAAT 1303 |
| QY | 206 | -----SerSerArgglyProThrLyAspGlyArgileLyasProaspValMet 221 |
| Db | 1304 | CAATATACATGCTTTCTAGAGGCCCAAGTCTGATGGAGCCCTCGTGTGAGCATCAGT 1363 |
| QY | 222 | AlaproglyThrPheIleLeuSerAlaArgSerSerLeuAlaProaspSerPheTrp 241 |
| Db | 1364 | GCACCAGGAGGCTGTTTCTGTGCTGCTAAC-----TGG 1399 |
| QY | 242 | AlaasnHisAspSerLysTyraLaTyMetGlyGlyThrSerMetAlaThrProIleVal 261 |
| Db | 1400 | ACATTGAGGGGAGCTCAG-----CTAATGAATGGGCATCAATGTCTCCCCCAATGCC 1453 |

COMMENT Contact: Screen SE
Entomology
University of Maryland
4112 Plant Sciences Building, College Park, MD 20742, USA.
FEATURES Location/Qualifiers
source 1..614

/organism="Metarhizium anisopliae"
/mol_type="mrna"
/strain="ARSEF 2575"
/db_xref="taxon:5530"
/clone="Ma#948"
/clone_lib="Metarhizium anisopliae ARSEF 2575"
/note="Vector: Unizap; Metarhizium anisopliae was grown on insect cuticle for 24 hours. A cDNA library was constructed in the unidirectional lambda vector, Unizap"

ORIGIN

Alignment Scores: 7.76e-08 Length: 614
Pred. No.: 182.50 Matches: 67
Score: 182.50
Percent Similarity: 42.68% Conservative: 38
Best Local Similarity: 27.24% Mismatches: 88
Query Match: 8.12% Indels: 53
DB: 9 Gaps: 12

US-09-985-689A-1-COPY (1-434) x AJ273402 (1-614)

| | | | |
|----|-----|---|-----|
| QY | 21 | GLYGLNGLYGLNILEVALALASPETHRGLYLEUASPETHRGLYARGASNASPSP | 40 |
| DB | 11 | GGTCAGGGTACTTGGCTATATATACATGACACTGGTATTGAG----- | 52 |
| QY | 41 | SERVETHISGLUAPHEARGGLYLSILETHRALALEUTHRALALEUGLYARGTHRA | 60 |
| DB | 53 | GCCTCCACCCCGAGTTTGGGTGCGGCACCTTTCTTAAGAGCTTCATCAGCGGTCA | 112 |
| QY | 61 | ASNALASNASPETHRASNGLYHISGLYTHRHISVALALAGLYSERVALLEUGLY | 80 |
| DB | 113 | AAC---ACTGATGCCACGCCATGGACTCACTGCGCTGGTACCATT----- | 160 |
| QY | 81 | SERTHRASNLYSGLYMETALAPROGLINALASNLEUVALPHEGLNLSERILEMET | 100 |
| DB | 161 | AGCAAGACCTTACCGTGTGTCACAAAGGCTAAGCTCTATGGTCAAGGTTCCTTG | 220 |
| QY | 101 | -----GLYGLYGLYLEUGLYGLYLEUPROSERASNLEUINLTHRLEUPHESER | 118 |
| DB | 221 | CAGGGCAGTGGTCTCTACTCCGTATCATCATGTCAGTGGCATGACTACGTTGC | 280 |
| QY | 119 | TYRSERALAGLY-----ALARGILEHISLTHRASNSETRTPGLYALALAV | 135 |
| DB | 281 | AAGACCGCGGCTGCCCAACGGCCGCTTCTTCATGAGCCTGGGA----- | 328 |
| QY | 136 | GLYALATYRTHRTHRASPERSERARGASNVALASPATYRVALARGLYASN | 155 |
| DB | 329 | GGTGGCTACTCGGCGTCCGTCACCAAGGTGCTGCTGCTTGGTCAATTCGTGTC | 388 |
| QY | 156 | ILELEUPHEALALAGLYASNGLUGLYPROASNGLYGLYTHRILESERALAPRO | 175 |
| DB | 389 | CTTCGCGTGGCGGTGGACGATACCGGGATGCCCAACACCTCT---CCCGCTTC | 445 |
| QY | 176 | ALALYSASNALALILETHRVALGLYALATHR-----GLUASNLEUARGPRO | 193 |
| DB | 446 | GAGCCTTCTGCTGCATCTGTGGTCTCTCGGAAATGACAGCCGATCTTCTCTTC | 505 |
| QY | 194 | SERTYRALASPASNILEASNHISVALALAGLNPHESERSERARGGLYPROTHR | 213 |
| DB | 506 | AACATAC----- | 511 |
| QY | 214 | GLYARGILELYSPROASPVALMETALAPROGLYTHRPHLEILEUSERALAARG | 233 |
| DB | 512 | GGCAGAGT---GTCGATATTTCTGCTCTGGTAGCAATGTCTTCTCCACC- | 559 |
| QY | 234 | LEUALAPROASPERSERPHETRIPALASNHISASPSERLYSTYRALATYR | 253 |

Db 560 -----TGGATTGGT-----GGCGGCACAAACACCATCTCTGGT 592
QY 254 ThrSerMetAlaThrPro 259
Db 593 ACCTNCATGGCTACTGCC 610

Search completed: March 15, 2004, 22:27:57
Job time : 2398 secs